

AI Risk Management Policy

How to govern, assess, approve, test, monitor and retire AI systems, from traditional machine-learning models to generative and agentic AI, built in-house or bought. Ready to adapt for your organisation.

Version 1.0 · October 2026 · Free template from Risk Frameworks Australia (riskfa.com) · No membership or sign-up

How to use this industry policy. Adopt the structure, terms, tiers and ratings as written, so that your organisation speaks the same language as the rest of the industry. Fill in the text in **[square brackets]** with your organisation's details, set the tiers, thresholds and timeframes to match your size, AI use and risk appetite, delete obligations that do not apply to you, and approve the Policy through your governance process. Regulatory requirements and guidance are summarised as at October 2026. Australia has no AI-specific law for the private sector as at that date, and the position is changing quickly: confirm the current requirements that apply to your organisation before relying on them. This template is general information only, provided free of charge, and is used at your own risk (see the disclaimer at the end).

Document control

Item	Detail
Document	AI Risk Management Policy
Organisation	[Organisation name]
Document owner	[Chief Risk Officer]
Approved by	[Board Risk Committee / Board]
Version / date approved	[1.0] / [date]
Review cycle	At least [annually], and after any material change in regulation, AI technology or the business, or a Critical or High AI incident
Parent policy	[Risk Management Framework / Operational Risk Management Policy]

1. Purpose

This Policy sets the minimum requirements for governing and managing the risks of **artificial intelligence (AI)** across [Organisation]. It ensures that:

- every AI use case is known, owned, recorded and assessed before it is used;
- the level of control matches the level of risk, so that low-risk uses are not held back and high-risk uses are tightly controlled;
- AI is designed, tested, approved, monitored, changed and retired in a controlled way;
- customers are treated fairly, are told when AI is used in decisions that affect them, and can challenge those decisions;
- personal information, confidential information and systems are protected from AI-specific threats;
- humans stay accountable for AI outcomes and can intervene or switch AI off; and
- the Board and management have accurate, timely information on AI use and AI risk, and AI use stays within risk appetite.

Australian financial services law is **technology neutral**: existing obligations apply to AI in the same way as to any other process or technology. ASIC has said that current licensee obligations, consumer protection laws and directors' duties apply to AI use, and that licensees should not take a wait-and-see approach to reform (ASIC REP 798, p 3 and p 34). APRA has said that its existing prudential standards apply to AI risk and that governance, risk management, assurance and operational resilience must keep pace with AI adoption (APRA letter to industry on AI, 30 April 2026).

The Policy supports [Organisation]'s obligations under its risk management framework, including, where they apply: APRA Prudential Standards CPS 220 *Risk Management*, CPS 230 *Operational Risk Management* and CPS 234 *Information Security*; the general obligations of AFS and credit licensees (*Corporations Act 2001* s912A; *National Consumer Credit Protection Act 2009* s47); the consumer protection provisions of the *Australian Securities and Investments Commission Act 2001*; the design and distribution obligations (*Corporations Act 2001* Part 7.8A); the *Privacy Act 1988*; Commonwealth, state and territory anti-discrimination laws; and the *Financial Accountability Regime Act 2023*.

2. Scope

This Policy applies to:

- all employees, contractors, directors and officers of [Organisation] and its subsidiaries;
- all AI systems and AI use cases, whether **built** in-house, **bought** or licensed from a

vendor, **embedded** in a vendor product or platform (for example AI features switched on in office, CRM or contact-centre software), or provided as a service by a third party;

- **traditional AI and machine learning** (for example credit scoring, fraud detection, pricing, claims triage, marketing propensity models), **generative AI** (for example chatbots, drafting, summarisation, code assistants) and **agentic AI** (AI that plans and takes actions in systems with some autonomy);
- AI used for internal purposes as well as AI that affects customers; and
- staff use of public or enterprise generative AI tools for work (section 15).

Statistical and rules-based models that are not AI (for example regulatory capital models, actuarial valuation models and deterministic business rules) remain covered by the **[Model Risk Management Standard]**. Where a model is both, both documents apply and the stricter requirement wins.

This Policy works with the RiskFA Incident & Issues Management Standard, the RiskFA Cyber Security Policy, the **[Privacy Policy]**, the **[Data Governance Policy]**, the **[Third-Party and Outsourcing Policy]** and the **[Change Management Standard]**. It does not repeat processes those documents cover.

3. Key definitions

Term	Meaning
AI system	A machine-based system that, for explicit or implicit objectives, infers from the input it receives how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments. AI systems vary in their levels of autonomy and adaptiveness after deployment. (OECD definition, as used in the terms and definitions of the Australian Government's <i>Guidance for AI Adoption</i> .)
Model	The mathematical or computational component of an AI system that has been trained or configured on data to produce outputs. One AI system may use several models, including third-party foundation models.
Traditional AI / machine learning	AI that learns patterns from historical data to classify, predict or score, for example a credit default or fraud model.
Generative AI	AI that generates new content, such as text, images, audio or code, in response to prompts. Includes large language models and the chatbots and assistants built on them.
Foundation model	A large, general-purpose model trained on broad data that can be adapted to many tasks, usually supplied by a third party.
Agentic AI (AI agent)	An AI system that can plan and carry out a sequence of actions towards a goal with limited human direction, for example by calling tools, accessing systems, sending messages or executing transactions.
AI use case	A specific business use of an AI system, in a specific context, for a specific purpose. The same AI system used for two purposes is two use cases, because the risk depends on how it is used.
High-risk use case	A use case rated High tier under section 8, for example one that makes or substantially informs decisions significantly affecting customers' rights or interests.
Prohibited use case	A use case [Organisation] will not undertake, set out in section 8.3.

Term	Meaning
Human oversight	The ability of competent, authorised people to understand, monitor, intervene in, override and stop an AI system, appropriate to the use case's tier. Common forms: human in the loop (a human approves each output before it takes effect), human on the loop (a human monitors and can intervene), and human in command (a human sets limits and can switch the system off).
Model risk	The risk of loss, customer harm or poor decisions from errors in the design, data, development, implementation or use of a model, or from using a model outside the purpose and conditions for which it was validated.
Drift	A change over time in the input data (data drift) or in the relationship between inputs and outcomes (concept drift) that reduces a model's accuracy or fairness.
Hallucination	A generative AI output that is presented as fact but is false, unsupported or invented.
Prompt injection	An attack in which crafted input, given directly or hidden in content the AI processes (such as a document, email or web page), causes the AI to ignore its instructions, disclose information or take unintended actions.
AI incident	An incident (as defined in the RiskFA Incident & Issues Management Standard) caused by, or involving, an AI system, including harmful, biased, false or unauthorised outputs or actions, and security or privacy incidents involving AI.
AI owner	The accountable executive for an AI use case (section 6).
Model owner	The person accountable for the technical performance of a model or AI system (section 6).
Explainability	The ability to describe, in terms suited to the audience, how an AI system reached an output and the main factors that drove it.
Contestability	The ability of a person affected by an AI-supported decision to challenge it and have it reviewed by a human.

Incident, near miss, issue, breach, root cause, action, customer remediation, critical operation, material service provider and reportable situation have the meanings in the RiskFA Incident & Issues Management Standard.

4. Principles

[Organisation]'s principles are consistent with Australia's AI Ethics Principles, the 6 essential practices in the Australian Government's *Guidance for AI Adoption* (October 2025) and the international standards referenced in section 23.

1. **Accountability.** A named person is accountable for every AI use case and its outcomes. Using AI never transfers or reduces accountability, including when a vendor supplies it.
2. **Risk-based.** Controls are proportionate to the use case's tier. The higher the potential harm, the stronger the controls, testing and oversight.
3. **Fairness.** AI must not unfairly discriminate against or exploit customers, including customers experiencing vulnerability. Fairness is tested before use and monitored after.
4. **Transparency.** People are told, in plain language, when they are interacting with AI, when AI-generated content is used and when AI makes or informs decisions that significantly affect them.
5. **Contestability.** Customers can challenge AI-supported decisions and have them reviewed by a person.
6. **Privacy.** Personal information is used in AI only lawfully, for proper purposes, and with the minimum data needed.
7. **Security.** AI systems, data, prompts and models are protected against AI-specific and conventional threats.
8. **Reliability.** AI systems are tested against clear acceptance criteria before use and monitored for drift, errors and hallucination throughout their life.
9. **Human oversight.** People can understand, intervene in, override and switch off AI systems. High-risk decisions require human involvement.
10. **Record keeping.** Decisions, assessments, testing and monitoring are documented so that they can be reviewed, audited and explained.

5. Three lines of accountability

Line	Who	Responsibilities
First line	Business units (AI owners), model owners, data and technology teams, procurement	Identify and register AI use cases; assess and tier them; design, build or buy, test, deploy, monitor and retire AI in line with this Policy; manage AI incidents and issues; train staff who use AI.
Second line	Risk (including model risk and operational risk), Compliance, Privacy, Information Security	Own this Policy and the tiering method; review and challenge tiering, impact assessments, testing and approvals; independently validate High tier models; monitor AI risk against appetite; report to management and the Board.
Third line	Internal Audit	Independently assess whether this Policy is designed and operating effectively, including the completeness of the AI register; raise audit issues, which are managed under the RiskFA Incident & Issues Management Standard.

6. Roles and responsibilities

Role	Responsibilities
Board	Sets the AI strategy and risk appetite for AI; maintains sufficient understanding of AI to provide effective challenge and oversight (APRA letter, 30 April 2026); approves Prohibited use case list changes; oversees material AI use, AI incidents and AI risk reporting.
Board Risk Committee	Approves this Policy; receives AI risk reporting at least [quarterly] (section 19); approves High tier use cases that are outside risk appetite or that involve a critical operation; is told of Critical AI incidents within [24 hours] .
AI Risk Committee (or [Executive Risk Committee])	Executive forum chaired by [the CRO or Chief Technology Officer] , with Risk, Compliance, Legal, Privacy, Information Security, Data, Technology and business members. Approves High tier use cases and material changes; reviews Medium tier approvals by sample; resolves tiering disputes; monitors the AI register, AI incidents and key risk indicators.
CEO and executive team	Ensure AI in their areas is used in line with this Policy and within risk appetite; act as AI owner, or appoint one, for each use case in their area.
Accountable persons (FAR, where applicable)	Consider AI use within their areas of responsibility when meeting their obligations under FAR Act s21. Where AI is used in a key function, the accountability map should make clear who is accountable for it (ASIC REP 798, question 4).
Chief Risk Officer (CRO)	Owns this Policy and the AI risk tiering method; ensures independent second line challenge and validation; reports AI risk to the Board Risk Committee.
Head of Compliance / General Counsel	Advise on legal and regulatory obligations for AI use, including consumer protection, design and distribution, anti-discrimination and intellectual property; assess AI incidents for reportability.
Chief Information Security Officer (CISO)	Sets and assures security controls for AI systems (section 13), including AI-specific threats, identity and access for AI agents, and security testing.

Role	Responsibilities
Privacy Officer	Ensures privacy impact assessments are done where required, advises on APP compliance including APP 1.7 to 1.9 transparency (section 14), and keeps the privacy policy current.
Chief Technology Officer / Chief Data Officer	Provide approved AI platforms and tools; maintain data governance, data quality and data lineage for AI; maintain technical records and logging.
AI owner	A senior manager accountable for an AI use case and its outcomes: registers and tiers it, ensures the impact assessment, testing and approvals are done, sets the human oversight model, monitors performance and customer outcomes, and decides on change or retirement.
Model owner	Accountable for the technical design, development or configuration, documentation, testing, performance monitoring and change control of the model or AI system. For bought AI, manages the vendor's technical information and updates.
Model validator	A person or team independent of the model's development, with the skills to assess AI systems. Validates High tier models before use and periodically (section 10.5).
All staff	Use only approved AI tools in line with section 15 and Appendix D; check AI outputs before relying on them; report AI incidents, near misses and concerns promptly.

Smaller organisations may give the AI Risk Committee's role to an existing executive risk committee, and combine the AI owner and model owner roles for bought, low-complexity tools. Independence of validation from development must be kept for High tier use cases.

7. AI use-case register

- Every AI use case must be recorded in the [AI register] **before** development, procurement or use starts, including pilots, proofs of concept and AI features enabled in existing vendor software. Appendix C sets the minimum fields.
- The register is the single source of truth for AI use at [Organisation]. It records AI tooling (platforms, models and products) and AI use cases, and links each to its owner, tier, risks, controls, vendors, approvals and incidents.
- AI owners keep their entries accurate and review them at least [annually] for Low tier, [half-yearly] for Medium tier and [quarterly] for High tier, and whenever the use case changes.
- Procurement, technology change, privacy impact assessment and third-party onboarding processes must check that any AI use is registered before they approve.
- Second line tests the register for completeness at least [annually], for example by comparing it with software inventories, vendor contracts, cloud AI service usage and network logs for public AI tools.

APRA expects regulated entities to maintain an inventory of AI tooling and AI use cases (APRA letter, 30 April 2026). ASIC asks licensees whether they have an AI inventory and are confident it is adequately maintained (REP 798, question 1). Guardrail 9 of the *Voluntary AI Safety Standard* includes maintaining an AI inventory.

8. Risk tiering of AI use cases

8.1 How to tier

The AI owner tiers each use case when it is registered, using the matrix in Appendix A, and second line reviews the tier. The tier sets the controls, approvals and oversight required (section 9). Tier on the **highest** criterion met. Re-tier whenever the use case, data, autonomy or vendor changes.

Tier	Summary	Typical examples
Prohibited	Uses [Organisation] will not undertake (section 8.3).	See section 8.3.

Tier	Summary	Typical examples
High	Makes or substantially informs decisions that significantly affect customers; acts with autonomy on customer accounts, payments or core systems; affects a critical operation; or uses sensitive information at scale.	Credit approval or limit decisions, insurance underwriting, pricing or claims denial, hardship or vulnerability triage, fraud blocking of customer transactions, financial advice content to customers, customer-facing generative AI that gives product information, AI agents that can change records or move money.
Medium	Customer-facing but with a human reviewing outputs before they take effect; or internal use that informs material business decisions or uses personal information.	Drafting customer letters reviewed by staff, complaints classification with human review, call summarisation used in case files, internal credit portfolio analytics, code assistants writing production code with mandatory review.
Low	Internal productivity use with no customer decision, no personal or confidential information beyond what is approved for the tool, and outputs checked by the user.	Summarising public documents, drafting internal emails, meeting notes in an approved enterprise tool, spell and grammar checking.

When in doubt between two tiers, choose the higher one. The tier is separate from the risk **rating** (Critical, High, Medium, Low) used for AI risks, incidents and issues, which follows the RiskFA Risk Management Framework and the RiskFA Incident & Issues Management Standard.

8.2 Factors that raise the tier

- **Customer impact:** the use case makes or substantially informs a decision that could significantly affect a person's rights or interests, such as access to credit, insurance, claims, superannuation benefits, pricing or account restrictions.
- **Vulnerable customers:** the use case targets, identifies or is likely to affect customers experiencing vulnerability.
- **Autonomy:** the AI acts without a human approving each action, or can act in production systems (agentic AI).
- **Data sensitivity:** the use case uses sensitive information (Privacy Act s6), health, biometric, children's, or large volumes of personal information, or confidential information.
- **Opacity:** the outputs cannot be explained to the level needed for the decision, or the vendor does not provide enough information to assess the model.
- **Scale and speed:** many customers or transactions affected before a human could notice an error.
- **Critical operations:** the use case supports a critical operation under CPS 230 or a material service provider arrangement.
- **Regulatory sensitivity:** the use case touches a regulated activity such as personal advice, credit assessment, design and distribution, AML/CTF, claims handling or complaints.

8.3 Prohibited use cases

[Organisation] must not use AI to:

- make a final decision to decline credit, insurance or a claim, close an account or deny a benefit **without** meaningful human review where required by section 11;
- exploit customers' vulnerabilities or behavioural biases, or manipulate customers into decisions against their interests;
- infer or use protected attributes (for example race, sex, disability, age, religion) to make decisions, except where the law expressly permits it (for example a lawful insurance exemption based on actuarial or statistical data) and Legal has approved it;
- create deceptive content, including deepfakes of real people, or impersonate a person, or present AI as a human when a customer asks;
- carry out biometric identification or emotion recognition of customers or staff without Legal and Privacy approval [or: at all] ;
- enter personal, sensitive or confidential information into public generative AI tools that

are not approved (section 15); or

- [any other use the Board decides is outside risk appetite] .

9. Controls by tier

Requirement	Low	Medium	High
Register entry	✓	✓	✓
AI impact assessment (customer, fairness, privacy, security, legal)	Short-form	Full	Full, reviewed by second line
Privacy impact assessment	If personal information is used	If personal information is used	Required
Security assessment	Approved-tool baseline	Required	Required, incl. adversarial testing
Testing against acceptance criteria	User acceptance	Documented testing	Documented testing, fairness testing and red-teaming (GenAI)
Independent validation	Not required	Sample by second line	Required before use and at least [annually]
Human oversight	User checks output	Human in the loop for customer outputs	Defined oversight model; human involvement in high- risk decisions
Customer transparency and contestability	Not usually needed	If customer-facing	Required (section 14)
Approval to deploy	AI owner	AI owner and second line	AI Risk Committee (Board Risk Committee where section 6 applies)
Monitoring and review	[Annual] review	[Quarterly] monitoring; [annual] review	Ongoing monitoring with thresholds; [quarterly] review

10. Lifecycle controls

These controls apply across the AI lifecycle. They reflect APRA's expectation of clear ownership and accountability "from design and development through to deployment, monitoring and decommissioning" (APRA letter, 30 April 2026), the lifecycle approach of ISO/IEC 42001 and ISO/IEC 23894, and the NIST AI RMF functions (Govern, Map, Measure, Manage).

10.1 Design and impact assessment

- Define the business purpose, the decision or task, the intended users, the people affected, and what success and failure look like.
- Consider whether AI is the right tool, and whether a simpler, more explainable method would do.
- Complete an AI impact assessment proportionate to the tier, covering customer outcomes, fairness, vulnerable customers, privacy, security, legal and regulatory obligations, explainability, human oversight, third-party dependence and operational resilience. ISO/IEC 42005:2025 gives guidance on AI system impact assessment.
- Engage the people who will use the system and, for High tier, those likely to be affected or their representatives (for example customer advocates), consistent with guardrail 10 of the *Voluntary AI Safety Standard*.
- Set measurable acceptance criteria for accuracy, fairness, robustness and, for generative AI, groundedness and harmful-output rates, before building or buying.

10.2 Data

- Use data that is fit for purpose, lawfully obtained and permitted for the use (section 14), with documented source, lineage and quality checks.
- Assess training and reference data for gaps, errors and historical bias, including under-representation of customer groups.
- Minimise personal information; de-identify where practical; never use production personal information in development or testing without approval.
- For retrieval-augmented generative AI, control which documents the AI can access and ensure access permissions follow the user's own permissions.
- Apply the [\[Data Governance Policy\]](#). APRA's CPG 235 *Managing Data Risk* is relevant guidance for APRA-regulated entities.

10.3 Development (built AI)

- Develop in controlled environments with version control for code, data, prompts, model parameters and configurations.
- Document the model or system: purpose, design, data, assumptions, limitations, known risks, testing results and intended use (for example a model card).
- Build in human oversight, logging, explainability and a way to switch off or fall back to a manual or non-AI process.

10.4 Procurement (bought AI)

Bought AI follows the same tiering and lifecycle as built AI. In addition, section 12 applies. Before buying, the AI owner and model owner must obtain enough information from the vendor to assess the model's purpose, training data sources, testing (including fairness and safety), known limitations, security, data handling and update practices. If the vendor cannot provide enough information for the tier, the use case must be re-tiered, redesigned or not used.

10.5 Validation and testing

- Test against the acceptance criteria before deployment, in conditions close to real use, and record the results.
- **Performance:** accuracy, stability and robustness, including on edge cases and stressed or unusual inputs.
- **Fairness and bias:** compare outcomes and error rates across customer groups, including groups protected by anti-discrimination law and vulnerable customers, using appropriate fairness measures and proxies where attributes are not held. Investigate and treat material differences before use.
- **Explainability:** confirm that outputs can be explained to the level the decision and the customer need (section 14).
- **Generative AI:** test for hallucination, groundedness in approved sources, harmful or inappropriate content, leakage of confidential information, and consistency. Carry out **red-teaming** (structured adversarial testing, including prompt injection and jailbreak attempts) for all customer-facing and High tier generative AI. NIST AI 600-1 *Generative AI Profile* describes generative AI risks and suggested actions.
- **Agentic AI:** test the limits on what the agent can do, its behaviour when tools fail or return unexpected results, its resistance to manipulation, and that it stops and escalates when it should.
- **Independent validation:** for High tier, a validator independent of development reviews the design, data, testing and limitations, and can require changes before approval.
- Testing evidence, issues found and their treatment are recorded. Failed criteria must be fixed, or the risk formally accepted under the RiskFA Incident & Issues Management Standard, before approval.

10.6 Approval and deployment

- Approvals follow section 9. The approver confirms that the impact assessment, testing, validation, privacy, security, vendor, transparency and oversight requirements are complete, using the checklist in Appendix B.
- Approval records the approved purpose, scope, conditions, limits and the monitoring thresholds. Use outside the approved purpose is a new use case.
- Deployment follows the **[Change Management Standard]**, with a tested rollback or switch-off and, for High tier, a staged release (for example a pilot or shadow run) where practical.

10.7 Monitoring

- AI owners monitor every deployed use case against the thresholds set at approval, including performance, **drift**, fairness, error and override rates, complaints, and, for generative AI, hallucination and harmful-output rates sampled by human review.
- Monitoring for High tier use cases is ongoing and automated where practical, with alerts to the model owner when a threshold is breached.
- A threshold breach is recorded and acted on: investigate, then retrain, recalibrate, restrict, add human review, or switch off. A breach that has caused or could cause harm is an AI incident (section 17).
- Human reviewers' overrides and customer complaints about AI outputs are analysed for patterns.

10.8 Change

- Material changes need re-assessment and re-approval at the use case's tier. Material changes include new purpose or users, new data sources, retraining, a new model or model version, changes to prompts or guardrails that affect outputs, increased autonomy, and vendor model updates.
- Contracts and monitoring must allow **[Organisation]** to know when a vendor changes or updates a model (section 12).
- Self-learning models that update in production are High tier unless second line agrees otherwise, and need limits and monitoring on what may change.

10.9 Periodic review and decommissioning

- Each use case is reviewed at the frequency in section 9 to confirm it is still needed, performing, fair, compliant and within its approved purpose.
- When an AI system is retired, the AI owner ensures a controlled transition (including to any replacement or manual process), removes access and integrations, deals with data, prompts and logs under the **[Records Management Policy]**, and updates the register.

11. Human oversight

- Each use case has a documented human oversight model (human in the loop, on the loop, or in command) suited to its tier and the decisions it supports.
- **High-risk decisions require human involvement and accountability** (APRA letter, 30 April 2026). For High tier use cases that make or substantially inform decisions to decline, restrict or price credit, insurance, claims or benefits, a competent person must be able to review and change the outcome before or promptly after it takes effect, and must do so where the customer asks (section 14).
- Human reviewers must have the time, training, information and authority to disagree with the AI. Monitoring of override rates helps detect rubber-stamping (automation bias).
- AI agents must operate within defined permissions, transaction and action limits, with logging of every action and a way to stop them immediately. Actions that are irreversible, high value or customer-affecting require human approval unless the AI Risk Committee approves otherwise.
- Every AI system has a documented fallback (manual process or alternative system) that has been tested, so that an AI system can be switched off without disrupting critical operations.

12. Third-party and vendor AI

[Organisation] remains responsible for outcomes when it uses AI provided by a third party. ASIC notes that licensees remain responsible for outsourced functions (REP 798, p 34). For APRA-regulated entities, CPS 230 requires the management of risks from service providers, including material service providers, and APRA expects entities to understand and manage AI supply-chain and concentration risk (APRA letter, 30 April 2026).

- Apply the **[Third-Party and Outsourcing Policy]** to every AI vendor, and assess whether the arrangement is with a **material service provider** under CPS 230 (for example where AI supports a critical operation). Material service provider arrangements must meet CPS 230, including APRA notification requirements.
- Map the AI supply chain, including foundation model providers, cloud platforms and other fourth parties that the vendor relies on, and record them in the register.
- Contracts for Medium and High tier use cases should cover: permitted use of **[Organisation]**'s data (including a prohibition on using it to train the vendor's models without approval), data location and retention, security standards, notification of incidents within **[24 hours]**, advance notice of material model changes and updates, information about model behaviour, testing and limitations, performance reporting, audit

and assurance rights, support for privacy and transparency obligations, and exit and transition assistance.

- Monitor vendor performance and model changes, and re-assess on material change.
- Assess concentration risk where several use cases depend on one provider, and maintain credible substitution, portability or exit arrangements for High tier and critical AI services.
- Check that AI features newly enabled by vendors in existing products are identified and registered before use.

13. Security of AI systems

AI systems are information assets and are protected under the RiskFA Cyber Security Policy and, for APRA-regulated entities, CPS 234. APRA has observed that AI adoption is materially changing the cyber threat landscape, including prompt injection, data leakage, insecure integrations and the manipulation or misuse of autonomous AI agents, and that identity and access management needs to adjust for non-human AI actors (APRA letter, 30 April 2026). ASIC has called for urgent uplift of cyber resilience fundamentals as frontier AI accelerates cyber threats (ASIC open letter, 8 May 2026).

- **Prompt injection and jailbreaks:** treat all content the AI processes as untrusted; separate instructions from data; filter inputs and outputs; limit what the AI can access and do.
- **Data leakage:** prevent confidential and personal information appearing in outputs to unauthorised users; apply data loss prevention to AI tools; restrict retrieval to what the user is permitted to see.
- **Model theft and tampering:** protect model files, weights, prompts, training data and configurations with access control, integrity checks and logging; protect against data poisoning.
- **AI agents and non-human identities:** give each agent its own identity with least-privilege access, privileged access management, time-limited credentials and full logging of actions.
- **AI-generated code:** apply the same code review, security testing and software composition analysis as to human-written code.
- **Security testing:** include AI-specific threats in penetration testing and red-teaming for Medium and High tier AI.
- **AI-enabled threats:** consider deepfake and AI-enabled social engineering in fraud, identity verification and payment controls.

14. Privacy, data and customer transparency

14.1 Privacy

The *Privacy Act 1988* and the Australian Privacy Principles (APPs) apply to personal information put into an AI system and to personal information it generates or infers. The OAIC's guidance on privacy and the use of commercially available AI products, and on privacy and developing and training generative AI models (both 21 October 2024), sets out how the APPs apply. In particular:

- generating or inferring personal information with AI is a collection and must comply with APP 3;
- using personal information in AI must be for the primary purpose of collection, or a permitted secondary purpose under APP 6;
- personal information used and produced must be accurate, up to date and complete (APP 10), which is relevant to hallucination;
- personal information must be protected (APP 11), including in vendor AI services;
- as a matter of best practice the OAIC recommends that organisations do not enter personal information, particularly sensitive information, into publicly available generative AI tools; and
- a privacy impact assessment is required for any AI use case that uses personal information in a new way, and for all High tier use cases.

14.2 Automated decisions in the privacy policy (from 10 December 2026)

From **10 December 2026**, APP 1.7 to 1.9 (inserted by the *Privacy and Other Legislation Amendment Act 2024*) require an APP entity's privacy policy to include information where it has arranged for a computer program to make, or do a thing substantially and directly related to making, a decision that could reasonably be expected to significantly affect an individual's rights or interests, using personal information about the individual. The privacy policy must describe the kinds of personal information used, the kinds of such decisions made solely by computer programs, and the kinds of decisions for which a computer program does a thing substantially and directly related to making the decision (APP 1.8).

- The OAIC's fact sheet (September 2026) says a computer program includes rule-based processes, AI and machine learning, and generative AI, and that a decision can be in scope even where a human reviews the program's output. The OAIC considers machine learning or generative AI outputs used to make such decisions generally in scope unless subject to extensive human oversight and control.
- The obligation generally stays with the entity making the decision, including when it

uses a vendor's program.

- The Privacy Officer must use the AI register to identify in-scope use cases and update the privacy policy by 10 December 2026, and keep it current as use cases change. The register records whether each use case is in scope (Appendix C).

14.3 Customer transparency, explanation and contestability

- Tell customers when they are interacting with an AI system (for example a chatbot), and when content they receive is substantially AI-generated, consistent with guardrail 6 of the *Voluntary AI Safety Standard*.
- Where AI makes or substantially informs a decision that significantly affects a customer, the customer must be told that AI was used, given the main reasons for the decision on request in plain language, and told how to ask for a human review.
- Customers must be able to reach a human. AI-supported decisions remain subject to the RiskFA Complaints Handling Policy, internal dispute resolution under ASIC RG 271 and external dispute resolution through AFCA.
- Statements to customers, the market and regulators about AI use, model performance and outputs must be accurate. ASIC notes that licensees must not make false or misleading representations, including about their AI use, model performance and outputs, and must ensure AI-generated representations are not false or misleading (REP 798, p 34; *ASIC Act* Part 2 Div 2).

14.4 Fairness, discrimination and product obligations

- Anti-discrimination laws (including the *Racial Discrimination Act 1975*, *Sex Discrimination Act 1984*, *Disability Discrimination Act 1992* and *Age Discrimination Act 2004*, and state and territory laws) apply to decisions made or informed by AI, including indirect discrimination through proxies.
- AFS and credit licensees must provide services efficiently, honestly and fairly. ASIC has said licensees should consider whether AI brings risks of unfairly biased or discriminatory treatment, or leaves them unable to explain outcomes, and must ensure AI is not used to act unconscionably or to exploit consumer vulnerabilities or behavioural biases (REP 798, p 34).
- Where AI is used in product design, target market determination, distribution, marketing or eligibility, the design and distribution obligations (*Corporations Act 2001* Part 7.8A; ASIC RG 274) continue to apply. AI-driven distribution must stay consistent with the target market determination.

15. Staff use of generative AI tools

- Staff may use for work only the generative AI tools listed as approved in the [Approved AI Tools List], within the data classifications approved for each tool.
- Public, consumer or personal-account generative AI tools must not be used with personal, sensitive, confidential, customer or [Organisation] proprietary information. APRA has raised concern about staff use of enterprise AI tools outside approved control frameworks (APRA letter, 30 April 2026).
- Staff remain responsible for any work they produce with AI. Outputs must be checked for accuracy, completeness and suitability before use, and must never be sent to customers without the review required for the use case.
- New business uses of an approved tool that go beyond the approved purposes must be registered as a use case.
- Appendix D sets the acceptable use rules, which all staff must acknowledge [annually]. Breaches are managed under the [Code of Conduct] and the RiskFA Consequence Management Framework.

16. Operational resilience

- AI that supports a critical operation must be included in the CPS 230 business continuity plan, with tested fallback processes and tolerance levels that take the AI dependency into account.
- Business continuity, backup and disaster recovery arrangements must cover AI systems, models, prompts and configurations (ASIC REP 798, question 9).
- Scenario analysis should include the loss of a key AI provider, widespread model failure, and AI-enabled cyber attacks.

17. AI incidents and issues

- AI incidents, near misses and issues are recorded, rated, escalated, notified, investigated and remediated under the **RiskFA Incident & Issues Management Standard**, using its Critical, High, Medium and Low ratings. They are flagged as AI-related in the register and linked to the use case.
- Examples of AI incidents: biased or wrong customer decisions; a chatbot giving false product information; personal information disclosed through an AI output; a successful prompt injection; an AI agent taking an unauthorised action; a vendor model change that degrades performance.
- The AI owner must be able to restrict or switch off an AI system immediately when it is causing or could cause harm, and must do so for any Critical or High AI incident unless the CRO agrees otherwise.
- Every AI incident is assessed for regulatory notification (for example CPS 230, CPS 234, ASIC reportable situations and Notifiable Data Breaches) whatever its rating, as the RiskFA Incident & Issues Management Standard requires.
- Customers harmed by AI outcomes are remediated under the RiskFA Incident & Issues Management Standard and ASIC RG 277.
- Root cause analysis for AI incidents should consider data, model design, testing gaps, monitoring thresholds, human oversight, vendor change and use outside the approved purpose.

18. Risk appetite and key risk indicators

The Board's the RiskFA Risk Appetite Statement should set appetite for AI risk, for example no appetite for AI decisions that unfairly discriminate, low appetite for unexplainable High tier decisions, and the conditions under which agentic AI may act without human approval. Key risk indicators measure whether AI use stays within appetite. Controls are tested for design and operating effectiveness under the **[Risk and Control Self-Assessment]** process and are adjusted when monitoring, incidents or the threat environment show they are not working.

Key risk indicator	Typical tolerance
AI use cases in use but not registered (found by discovery)	[0]
High tier use cases without current independent validation	[0]
Use cases overdue for periodic review	≤ [5%] ; no High tier more than [30 days] overdue

Key risk indicator	Typical tolerance
Monitoring threshold breaches (performance, drift, fairness) not actioned within [10 business days]	[0]
Generative AI hallucination or harmful-output rate in sampled customer-facing outputs	$\leq$ [x%], set per use case at approval
Human override rate outside expected range	Investigate outside [range set at approval]
AI-related complaints	Trend monitored; investigate any rise of more than [20%] quarter on quarter
AI incidents rated High or Critical	[0]; any is reported to the Board Risk Committee
Blocked attempts to use unapproved public AI tools	Trending down
Staff who have completed AI acceptable use training	$\geq$ [95%]
High tier use cases relying on a single AI provider without a tested exit or substitution plan	[0]

These indicators and tolerances are common practice, not regulatory requirements. Set your own in line with your risk appetite.

19. Reporting

Second line reports on AI risk to the [AI Risk Committee] [monthly] and to the Board Risk Committee [quarterly]. Reports cover at least:

- the number of AI use cases by tier, status and business unit, and new High tier approvals;
- key risk indicators against tolerance (section 18);
- AI incidents, issues and customer remediation;
- validation and monitoring results, including fairness findings;
- third-party AI concentration and material vendor changes; and
- regulatory developments and any changes needed to this Policy.

20. Training and capability

- All staff complete AI acceptable use training at induction and [annually] .
- AI owners, model owners, human reviewers and approvers complete role-specific training on this Policy, AI risks, limitations and their responsibilities before taking up the role.
- Directors and executives receive AI briefings at least [annually] so they can provide effective challenge, as APRA expects of boards.
- Second and third line staff who review AI must have the technical capability and tools to assess AI systems, including probabilistic models and agentic workflows (APRA letter, 30 April 2026), using specialists where needed.

21. Assurance

- Second line reviews a sample of Low and Medium tier registrations and approvals at least [half-yearly] .
- Internal Audit reviews the design and effectiveness of this Policy, and the completeness of the AI register, at least every [2 years] , and audits High tier use cases on a risk basis.
- [Organisation] may seek certification or independent assessment against ISO/IEC 42001:2023 *Artificial intelligence management system*. Certification is voluntary.

22. Compliance with this Policy and exemptions

Using AI that has not been registered, tiered or approved under this Policy, or using AI outside its approved purpose, is a breach of this Policy and is managed under the RiskFA Incident & Issues Management Standard and the RiskFA Consequence Management Framework, and considered in performance and remuneration outcomes under the RiskFA Remuneration Policy (CPS 511 where it applies). Exemptions need the written approval of the CRO, must be time-limited, and are reported to the Board Risk Committee. No exemption may permit a Prohibited use case.

23. Regulatory and standards context (as at October 2026)

Source	Status	What it means for this Policy
Australian Government <i>Voluntary AI Safety Standard</i> (DISR, September 2024): 10 guardrails	Voluntary	Basis for the accountability, risk management, data, testing, human oversight, transparency, contestability, supply chain, record keeping and stakeholder controls in this Policy.
<i>Guidance for AI Adoption</i> (National AI Centre, 21 October 2025; documents updated May 2026, now published on ai.gov.au): 6 essential practices: decide who is accountable; understand impacts and plan accordingly; measure and manage risks; share essential information; test and monitor; maintain human control	Voluntary. The Government describes it as updated and simplified guidance that evolves the Voluntary AI Safety Standard	Sections 4 to 11 are consistent with the 6 practices.
Mandatory guardrails for AI in high-risk settings (proposals paper, September 2024)	Not proceeding. The Government "will not proceed at this time" with the proposals (DISR consultation outcome)	No AI-specific law applies to [Organisation] as at October 2026. Existing laws apply.
<i>National AI Plan</i> (2 December 2025)	Government policy	Regulation of AI builds on existing, largely technology-neutral laws, with regulators responsible in their domains, targeted laws where needed, and an Australian AI Safety Institute to monitor and advise.

Source	Status	What it means for this Policy
APRA letter to industry on AI (30 April 2026)	Supervisory expectations. APRA states that its existing prudential standards apply to AI risk	Board AI literacy and oversight; AI frameworks (policy, standards, guidance); lifecycle ownership; AI inventory; human involvement for high-risk decisions; staff training; AI security; supply-chain and concentration risk; continuous, integrated assurance. APRA has said it will take stronger supervisory action and, where appropriate, pursue enforcement where entities fail to adequately identify, manage or control AI risks.
APRA and ASIC, <i>Resilience at Frontier AI Speed</i> (27 August 2026) and ASIC open letter (8 May 2026)	Regulator information paper and open letter	Cyber resilience fundamentals, board preparedness, third-party concentration and defensive use of AI.
ASIC REP 798 <i>Beware the gap</i> (29 October 2024)	Regulator report and questions for licensees	Existing licensee obligations, consumer protection laws and directors' duties apply to AI; AI governance questions for licensees across 11 topics (pp 35–36).
<i>Privacy Act 1988</i> , APP 1.7 to 1.9 (from 10 December 2026); OAIC AI guidance (21 October 2024)	Law (APPs); OAIC guidance explains how the law applies	Privacy policy disclosure of automated decisions; APP 3, 6, 10 and 11 apply to AI inputs and outputs.
<i>Corporations Act 2001</i> s912A, Part 7.8A; <i>NCCP Act</i> s47; <i>ASIC Act</i> Part 2 Div 2; anti-discrimination laws; <i>FAR Act</i>	Law	Efficiently, honestly and fairly; adequate resources and risk management; design and distribution; no misleading or unconscionable conduct; no unlawful discrimination; accountability.
ISO/IEC 42001:2023; ISO/IEC 23894:2023; ISO/IEC 42005:2025	Voluntary international standards	AI management system (certifiable); AI risk management guidance; AI system impact assessment guidance.
NIST AI RMF 1.0 (AI 100-1, January 2023) and NIST AI 600-1 <i>Generative AI Profile</i> (July 2024)	Voluntary (United States)	Govern, Map, Measure, Manage functions; generative AI risks and actions. NIST has said AI RMF 1.0 is being revised.

This summary is general information and may not be complete for your organisation. It does not cover overseas laws such as the EU AI Act, which may apply if you offer AI-enabled services in other jurisdictions. Confirm current requirements with the regulator's guidance or your legal adviser.

24. Related documents

- the RiskFA Risk Management Framework, the RiskFA Risk Appetite Statement and the RiskFA Operational Risk Management Policy
- RiskFA Incident & Issues Management Standard
- the RiskFA Cyber Security Policy, [\[Privacy Policy\]](#) , [\[Data Governance Policy\]](#) and [\[Records Management Policy\]](#)
- [\[Third-Party and Outsourcing Policy\]](#) , [\[Change Management Standard\]](#) and [\[Model Risk Management Standard\]](#)
- the RiskFA Business Continuity Policy, the RiskFA Complaints Handling Policy, [\[Code of Conduct\]](#) , the RiskFA Consequence Management Framework and the RiskFA Remuneration Policy
- APRA: letter to industry on AI (30 April 2026); CPS 220, CPS 230 and CPG 230, CPS 234 and CPG 234, CPG 235 *Managing Data Risk*
- ASIC: REP 798 *Beware the gap: Governance arrangements in the face of AI innovation*; RG 271 *Internal dispute resolution*; RG 274 *Product design and distribution obligations*; RG 277 *Consumer remediation*
- OAIC: *Guidance on privacy and the use of commercially available AI products*; *Guidance on privacy and developing and training generative AI models*; APP 1.7–1.9 Transparency Obligation Fact Sheet; APP Guidelines Chapter 1
- Department of Industry, Science and Resources: *Voluntary AI Safety Standard*; *Guidance for AI Adoption*; *National AI Plan*
- Standards: ISO/IEC 42001:2023, ISO/IEC 23894:2023, ISO/IEC 42005:2025, AS ISO 31000 *Risk management*; NIST AI RMF 1.0 and NIST AI 600-1

Appendix A. AI use-case risk tiering matrix

Answer each question. The use case takes the **highest** tier reached. Any Prohibited answer stops the use case.

Criterion	Low	Medium	High	Prohibited
-----------	-----	--------	------	------------

Criterion	Low	Medium	High	Prohibited
Customer decisions	No customer decision	Informs a customer outcome; a human reviews before it takes effect	Makes or substantially informs a decision significantly affecting a customer (credit, insurance, claims, pricing, benefits, account restrictions)	Final adverse decision with no human review where section 11 requires it
Customer interaction	None	Customer receives AI-drafted content after human review	Customer interacts directly with AI, or receives AI content without human review	AI presented as a human, or deceptive content
Vulnerable customers	Not affected	May be affected indirectly	Targets, identifies or materially affects customers experiencing vulnerability	Exploits vulnerability or behavioural biases
Autonomy	User checks every output	Human in the loop	Acts without prior human approval, or AI agent with access to production systems	Irreversible high-value actions with no human control
Data sensitivity	Public or internal information approved for the tool	Personal or confidential information in an approved enterprise tool	Sensitive information, large volumes of personal information, or children's information	Personal or confidential information in an unapproved public tool
Explainability	Not needed	Outputs can be explained	Limited explainability for a significant decision	Cannot be explained where an explanation is legally required
Operations	No operational dependency	Supports a business process with a workaround	Supports a critical operation or a material service provider arrangement	n/a

Criterion	Low	Medium	High	Prohibited
Protected attributes	Not used	Not used	Used only where lawful and approved by Legal	Used to discriminate unlawfully

Appendix B. Pre-deployment checklist

- Use case registered, with AI owner, model owner and tier recorded; tier reviewed by second line.
- AI impact assessment completed and reviewed at the level required for the tier.
- Privacy impact assessment completed where required; APP 1.7–1.9 scope decided and the privacy policy updated if in scope.
- Legal and Compliance review completed for customer-facing, advice, credit, insurance, claims, marketing or distribution use cases.
- Data sources, lineage, quality and permitted use documented; bias in data assessed.
- Security assessment completed, including AI-specific threats; identity and access set up for any AI agent.
- Testing against acceptance criteria completed and recorded, including fairness testing and, for generative AI, hallucination testing and red-teaming.
- Independent validation completed (High tier) and findings resolved or formally accepted.
- Vendor due diligence and contract terms completed; material service provider assessment done; supply chain recorded.
- Human oversight model documented; reviewers trained; fallback and switch-off tested.
- Customer disclosures, explanations and contestability process in place where required.
- Monitoring metrics, thresholds, alerts and owners set; review date set.
- Business continuity and rollback arrangements in place.
- Approval obtained at the level required by section 9, with conditions recorded.

Appendix C. AI register fields

Field	Notes
Unique ID, use case name and description	One entry per use case
Business purpose and decision or task supported	Including the approved purpose and limits
AI type	Traditional machine learning, generative AI, agentic AI, or other
Built, bought, embedded or service	Including vendor name, product, model and version, and foundation model provider
Business unit, process and critical operation (if any)	
AI owner and model owner	Named people
Tier and rationale; tier history	Using Appendix A
Customers affected (number, type, vulnerability)	
Personal information used; sensitive information; APP 1.7–1.9 in scope (yes / no) and reason	
Human oversight model and fallback	
Impact assessment, privacy impact assessment and security assessment references	
Testing, fairness testing, red-teaming and validation results and dates	
Approval, approver, date and conditions	
Monitoring metrics, thresholds and latest results	
Material service provider (yes / no); contract reference; fourth parties	
Linked risks, controls, incidents, issues and complaints	
Status (proposed, pilot, live, suspended, retired) and next review date	

Appendix D. Acceptable use rules for staff

1. **Use approved tools only.** Use only the AI tools on the [\[Approved AI Tools List\]](#), through your [\[Organisation\]](#) account.
2. **Protect information.** Never enter personal, sensitive, customer, confidential or [\[Organisation\]](#) proprietary information into a tool not approved for that information. If unsure, do not enter it.
3. **You are responsible.** Check AI outputs for accuracy, completeness, bias and suitability. AI can be confidently wrong.
4. **Customers come first.** Do not send AI-generated content to customers, or use it in a customer decision, unless the use case is approved and the required review is done.
5. **Be open.** Do not present AI-generated work as checked when it is not. Follow the rules on disclosing AI use to customers.
6. **Respect others' rights.** Do not use AI to create content that infringes copyright, impersonates a real person or misleads anyone.
7. **Watch for manipulation.** Treat instructions inside documents, emails or web pages processed by AI with suspicion, and be alert to deepfake voices, videos and messages, especially in payment or access requests.
8. **Register new uses.** Any new business use of AI, including AI features turned on in existing software, must be registered before use.
9. **Report problems.** Report AI incidents, near misses, unexpected outputs and suspected misuse to [\[your manager and the incident register\]](#) straight away.
10. **When in doubt, ask** [\[the AI Risk team / Risk and Compliance\]](#).



[Risk Taxonomy](#) [Operations](#) [Library](#) [Obligations Register](#) [Why a standard](#) [About](#)

Disclaimer

© 2026 Risk Frameworks Australia

Free general information only, provided free of charge. Not legal, financial or other professional advice, and used at your own risk. [Read the full disclaimer.](#)