# Technical Paper: Considerations for AI adoption - Applying the AI6
August 2026

## Contents

## 1. Status of Technical Paper

This Technical Paper has been prepared by the Data Science and Artificial Intelligence Practice Committee (DSAIPC) of the Actuaries Institute, for the purpose of assisting Members involved in design, development, deployment, oversight or assessment of AI systems, and to contribute to AI risk conversations.

This Technical Paper does not represent a Professional Standard or Practice Guideline of the Institute. The information contained in this Technical Paper is commentary and general information only. It is not mandatory for Members to consider this Technical Paper in their work. This Technical Paper does not constitute legal advice. Any interpretation or commentary regarding specific legislative or regulatory requirements reflects the expectations of the Institute but does not guarantee compliance under applicable legislation or regulations. Accordingly, Members should seek clarification from the relevant regulator and/or seek legal advice if they are unsure or require specific guidance regarding their legal or regulatory obligations.

In all cases, where conflict may exist between this Technical Paper and either the Actuaries' Code or a Professional Standard, the Actuaries' Code or the Professional Standard takes precedence.

This Technical Paper replaces the Technical Paper on Automated Decision-Making Systems (October 2020, reclassified January 2023), which has been retired in light of the substantial development of AI governance practice, frameworks, and Australian guidance since its publication.

Ongoing feedback from Members is encouraged; any feedback should be directed to the DSAIPC via ppd@actuaries.asn.au

## 2. Introduction and Scope

AI is changing how organisations operate across every sector, and the actuarial profession is no exception. AI systems are increasingly embedded in pricing, reserving, claims processing, fraud detection, customer service and other functions, with decisions that may affect people's financial wellbeing and access to essential products and services.

The landscape of AI governance frameworks is crowded. ISO/IEC 42001[1], the NIST AI Risk Management Framework[2], the OECD's AI principles[3], and now Australia's own Guidance for AI Adoption[4] (sometimes called the "AI6") all sit alongside one another. The volume of guidance can make it difficult for practitioners to know where to start. Members are well placed to contribute to AI risk conversations (with the Actuarial Control Cycle providing a natural foundation for structured AI governance) but may lack a structured way to do so in their day-to-day work.

This Technical Paper takes the AI6 as its anchor and considers, given the framework exists, what Members might usefully do with it. It suggests using these as a set of principles, good practices and illustrative examples that Members may find useful in considering the design, development, deployment, oversight or assessment of AI systems. The paper is likely to be most useful to Members who are regularly involved in these activities. Appointed Actuaries may also find this paper useful as context, since many organisations have either started pilots involving AI systems, or deployed AI systems. It may be relevant to include commentary on AI in the Financial Condition Report, or other reports and documents.

We recommend reading this Technical Paper alongside the AI6. The AI6 sets the foundation by outlining practices that serve as a mix of good governance and risk management for AI. The questions posed in this paper support Members to contextualise how those practices might apply to their organisation.

Members work in different organisations with different risk frameworks. This Technical Paper does not prescribe a single approach. It is a matter of organisational context and professional judgement to determine in which circumstances this Technical Paper should be considered relevant, noting it forms part of the Professional Governance Material available to a Member.

### 2.1. The AI risk landscape

The AI6 sets out how to govern AI systems. In this Technical Paper we highlight some of the risks being governed. A risk that may be of particular relevance is that practitioners apply governance practices without engaging with the substance behind them. Before turning to principles and good practices, it may be helpful to be explicit about the risk landscape.

It is common to hear AI categorised into three types: narrow, generative and agentic. These categories are not always mutually exclusive, but each presents different governance considerations.

[1] https://www.iso.org/standard/42001
[2] https://www.nist.gov/itl/ai-risk-management-framework
[3] https://www.oecd.org/en/topics/sub-issues/ai-principles.html
[4] https://www.ai.gov.au/staying-safe-and-responsible/essential-ai-practices

### 2.1.1. Narrow AI

Narrow AI refers to AI systems designed to perform a specific set of tasks. Examples include an XGBoost classification model, a neural network for image recognition, or a generative AI application tailored to a narrow scope through specialised prompts and workflow design.

In these systems, bias in data, model drift, and poor performance on edge cases are longstanding concerns. Transparency and explainability may present a more novel challenge. Unlike traditional actuarial models such as Generalised Linear Models (GLMs), where parameters are interpretable and logic is traceable, many narrow AI models are effectively black boxes whose complexity makes meaningful explanation difficult. A decision that cannot be explained may be difficult to defend to a regulator, or to a customer who believes they have been adversely affected.

### 2.1.2. Generative AI

Generative AI refers to systems that produce new content such as text, images, audio, video or code. Generative AI introduces a further set of less familiar concerns. Explainability is even more limited than with narrow AI, with outputs generated through processes that are opaque even to their developers. Techniques such as chain-of-thought prompting can make reasoning steps more visible, but do not resolve the underlying opacity. This remains an active area of research: the emerging field of mechanistic interpretability aims to reverse-engineer the internal mechanisms of these systems.

Generative AI also carries additional risks. It can *hallucinate* plausible-sounding but fabricated information, and sensitive data may be exposed when users input confidential information into third-party generative AI systems where data may be retained or accessed beyond its intended purpose. Generative AI also exposes organisations to *adversarial* risks including (but not limited to) prompt injection (malicious instructions smuggled into inputs the model processes) and data poisoning (training data manipulated to bias model behaviour).

### 2.1.3. Agentic AI

Agentic AI compounds these concerns. To delineate from the categories above, here we are considering an agentic system which can take multiple consequential steps autonomously based on high-level inputs or instructions: for example, sending communications, updating records, triggering transactions. Errors may propagate within such systems and cause harm before any human review. This may make meaningful human oversight both more important and more difficult to maintain.

Taken together, these AI characteristics suggest that existing governance approaches may not be sufficient on their own. Frameworks designed for traditional model risk management capture some but not all of what Members may now need to address.

## 2.2. Existing frameworks and the AI lifecycle

AI6 is one of several frameworks a Member may encounter. Others include the NIST AI Risk Management Framework (a peak US reference developed in cyber and security), ISO/IEC 42001 (the international AI management standard with which AI6 broadly aligns), and Australian work on responsible AI practice from organisations including CSIRO. The point is not that one framework is correct and others are wrong. What matters more is that an organisation chooses a framework (which may be a hybrid approach), applies it consistently, and uses it to drive substantive risk conversations.

It may be helpful to contextualise these governance practices within the AI lifecycle. For actuaries, the concept should feel familiar. A useful approach is to articulate the lifecycle as a sequence of stages: for example, problem definition, data collection, exploratory data analysis, modelling, evaluation, deployment, use, and retirement, each with its own ethical checkpoints. Such a lifecycle maps naturally onto the Actuarial Control Cycle, with early stages defining the problem, middle stages designing the solution, and the final stage monitoring results. This Technical Paper does not prescribe a specific lifecycle. The relevant consideration is that a Member works within a defined lifecycle and can articulate where each governance practice applies.

## 2.3. Voluntary Guidance

The AI6 guidance is voluntary. Australia has no AI-specific legislation, although AI use is not unregulated: existing privacy, anti-discrimination, consumer protection, financial services and professional standards laws continue to apply to AI systems as they do to other tools.

When the previous voluntary standard (the Voluntary AI Safety Standard, VAISS) was released alongside proposed mandatory guardrails in September 2024, the strong linkage created the impression that all elements of the VAISS would effectively become mandated. The Actuaries Institute submitted feedback on the proposed mandatory guardrails, observing that the VAISS was not written in a way that could be moved to a regulatory mandate line-by-line.

The framing of AI6 as guidance is a step forward. Organisations may choose to adopt practices that make sense in their context, and adapt or set aside those which do not. AI6 should not be conflated with legal requirements. However, implementing these practices may also position organisations well for whatever mandatory framework may emerge.

For Members in APRA-regulated entities, the practical position turns less on AI-specific legislation than on how existing prudential frameworks apply to AI. In its April 2026 Letter to Industry on Artificial Intelligence[5], APRA confirmed that its prudential framework is technology- and vendor-agnostic, and that it expects regulated entities to manage AI risk through existing standards on governance, operational resilience, information security and third-party arrangements. APRA also signalled that supervisory attention will follow where AI risks are not managed in proportion to an entity's size, scale and complexity, with enforcement available where warranted. The position in superannuation and banking is consistent.

Further, ASIC's May 2026 Open letter to AFS licensees and market participants[6] called for an uplift in cyber resilience due to AI accelerating the capability, accessibility, speed and scale of cyber threats. The recommendations primarily related to information security and governance best practices, but are also relevant to the secure design, development and deployment of AI systems.

The broader point: voluntary frameworks such as AI6 do not displace existing obligations, and supervisory expectations on AI may emerge not through new AI-specific law but through existing prudential, conduct and professional frameworks.

## 3. The AI6 Framework

In early 2026, the National AI Centre[7] (NAIC) released updated guidance consolidating previous frameworks into six core practices for responsible AI adoption (the "AI6"). The Actuaries Institute supports the release, noting that the guidance enables leaders across sectors to develop a shared fluency in AI governance and strengthen collaboration.

The guidance comes in two formats: a Foundations document[8] for organisations getting started, and an Implementation Practices guide[9] broadly aligned with ISO/IEC 42001. The tiered approach recognises that organisations are at different stages of AI maturity. A third guide, Be Clear About AI -generated Content [10], considers when and how to disclose use of AI.

The DSAIPC views the AI6 guidance as offering a useful set of practical guidelines for governing AI. The AI6 is accessible, practical, free to access, and written for organisational leaders rather than technical specialists. It may help productive conversations between Members, executives, data scientists, engineers, and compliance teams. The international alignment may also position organisations well for global operations.

---

[5] https://www.apra.gov.au/news-and-publications/apra-letter-industry-artificial-intelligence-ai
[6] https://download.asic.gov.au/media/xhrf1w0e/26-092mr-open-letter-to-afs-licensees-and-market-participants.pdf
[7] https://www.ai.gov.au/
[8] https://www.ai.gov.au/staying-safe-and-responsible/essential-ai-practices/guidance-ai-adoption-foundations
[9] https://www.ai.gov.au/staying-safe-and-responsible/essential-ai-practices/guidance-ai-adoption-implementation-guidance
[10] https://www.ai.gov.au/staying-safe-and-responsible/essential-ai-practices/be-clear-about-ai-use

### 3.1. The Six Practices

The six practices, summarised here, are:

- **Decide who is accountable.** Establish clear governance structures with specific accountability for AI systems throughout their lifecycle.
- **Understand impacts and plan accordingly.** Identify and manage AI's effects on stakeholders, with particular attention to vulnerable groups, and provide channels to contest decisions.
- **Measure and manage risks.** Apply AI-specific, context-dependent risk management with controls proportionate to risk level.
- **Share essential information.** Be transparent about AI use and capabilities, including through AI registers and disclosure to affected people.
- **Test and monitor.** Conduct rigorous pre-deployment testing and ongoing post-deployment monitoring.
- **Maintain human control.** Ensure meaningful human oversight matched to the system's autonomy and the stakes involved.

These practices represent high-level guidance. The rest of this Technical Paper considers good practices that Members may wish to consider when applying them.

## 4. Good practices for Members

In this section we set out good practices that Members may wish to consider when applying AI6 in actuarial work. The practices are organised around key themes inspired by the AI6, where actuarial expertise may be most valuable, and where Members may be best placed to drive substantive conversation.

Like the practices in any framework of this kind, the items below are not exhaustive, nor will they all be relevant or appropriate in every situation. They should not be used as a checklist. Good practice in one context may be excessive or inadequate in another; a Member exercises professional judgement on what is appropriate in the situation at hand.

Each subsection is structured around: context (why the AI6 theme matters), examples of risks, relevance for actuarial practice, and questions a Member may consider asking. Where useful, we note the stage of the Actuarial Control Cycle (problem definition, solution design, monitoring) to which the practice most closely relates.

## 4.1. Risk Management

### 4.1.1. Context

AI systems may behave differently depending on the context in which they are deployed. The same model can be appropriate for one use case and inappropriate for another. Without context-dependent risk screening, an organisation may either over-govern low-risk applications (slowing useful work) or under-govern high-risk ones (accepting potential harm to customers and to the organisation). This consideration is most relevant at the problem definition and solution design stages of the Actuarial Control Cycle, with ongoing relevance during monitoring.

### 4.1.2. Examples of Risk

- A pricing process focussing on whole-of-portfolio outcomes may leave undetected undesirable, unstable or unfair outcomes for small segments of the population.
- Generative AI applications connected to an organisation's broader IT environment that are exposed to potential adversarial manipulation through prompt injection or data poisoning. This may be an issue even where the intended use case appears low-risk, for example, summarising customer-uploaded documents.
- Model drift, where a system that worked well in stable conditions starts making poor decisions during a shock event such as a natural disaster, a pandemic, or a sudden change in consumer behaviour.

### 4.1.3. Relevance for Actuarial Practice

AI6 calls for context-dependent risk screening with proportionate controls. Risk assessment is part of the core actuarial skillset, and Members may be well positioned to support its effective execution. Members may work across both traditional and generative AI systems, each with a different risk profile. Members should consider engaging with broader subject matter experts (spanning business and technology stakeholders) to define the key risks associated with the AI system, the potential impacts (frequency and severity), and the appropriate controls to bring residual risk to an acceptable level.

### 4.1.4. Questions a Member may consider asking

- What criteria distinguish high-risk and low-risk use cases, and how do our practices vary accordingly? This could be considered as part of a risk appetite process. For example, decisions that significantly impact individuals' rights, that specifically target vulnerable populations, or affect large swathes of the population might be deemed high-risk.
- Are risk assessments conducted at a use case or system level, rather than at a model level? The evaluation only makes sense when considering the end impact on customers or users, as the same model deployed in different contexts might result in completely different risk profiles.

- What is the incident-response process for AI failures, and how does it vary based on the risk profile of the system? For example, high-risk systems are likely to warrant tighter thresholds for alerting and escalation, as well as faster service level agreements for incident response and resolution.

## 4.2. System Evaluation

### 4.2.1. Context

AI systems may not be stable in the way some traditional actuarial models are. They can drift as the underlying environment changes, behave unpredictably on edge cases, and potentially be manipulated when exposed to user inputs. Pre-deployment validation alone may not be sufficient. This consideration spans the solution design and monitoring stages of the Actuarial Control Cycle.

### 4.2.2. Examples of Risk

- A claims-management AI that drifts unevenly, maintaining average accuracy but producing materially worse outcomes for specific demographic or geographic segments, visible only in detailed monitoring data, not in headline metrics.
- AI red-teaming[11] gaps: a system that has never been adversarially tested may have unknown vulnerabilities for a motivated attacker to exploit.
- Reserving models may have poor accuracy but this may not become clear until years after deployment, the period over which the model is being relied upon.

### 4.2.3. Relevance for Actuarial Practice

AI6 recommendations on pre-deployment testing and post-deployment monitoring align with actuarial expertise in model validation and the broader control cycle. Testing AI systems may require evaluating both the happy path (performance under normal conditions or intended use) and potential failure modes (edge cases, adversarial attacks). A model that performs well on average may behave poorly on extreme outcomes.

As generative AI spreads, adversarial considerations may become more relevant. Even seemingly low-risk applications (such as processing uploaded customer files, or conducting research on the internet) can expose organisations to attack, particularly if such attacks can impact an organisation's broader IT estate. Good practice may include conducting AI red-teaming and broader cybersecurity assessment on AI systems, calibrated to the risk profile of the use case, to assess robustness to exploits.

[11] For instance, refer https://www.paloaltonetworks.com.au/cyberpedia/what-is-ai-red-teaming

AI systems can also degrade in ways that traditional, sample-based or point-in-time assurance may not detect. Where a model learns or adapts, continuous validation may be more relevant than periodic sign-off. Thresholds and alerts should be calibrated to the use case's risk profile. This is now reflected in supervisory expectations.

#### 4.2.4. Questions a Member may consider asking

- Does the testing regime evaluate both overall system performance and behaviour on edge cases? What are the key metrics at individual model level and at overall system level? What is the acceptable error rate, and how does this compare to the status quo?
- What adversarial testing has been conducted to assess the robustness of the system? Which stakeholders are best placed to inform the design of the threat model and attack vectors? What conditions might warrant independent or external testing?
- What post-deployment monitoring is in place? What thresholds and alerts trigger human review, model retraining, or system retirement? How often is the system periodically reviewed to determine whether it remains fit for purpose?

### 4.3. Supply Chain Considerations

#### 4.3.1. Context

AI6's principle on sharing essential information covers transparency in two directions: outward, to customers, regulators, and other stakeholders affected by AI use; and inward, along the AI supply chain, so that the organisation understands the systems it is deploying. Both are important. We focus here on the supply chain dimension, where Members may most often encounter material gaps. This consideration is most relevant at the solution design and monitoring stages of the Actuarial Control Cycle, with implications for problem definition where a system has not yet been chosen.

#### 4.3.2. Examples of Risk

- Forced upgrade cycles, whether due to vendors choosing to sunset a particular model version, or silently updating their underlying model, undermining reliance on point-in-time validation.
- Service degradation or discontinuation by a third-party provider, with no internal fallback path.
- Concentration risk where multiple AI use cases depend on a single vendor, foundation model or platform, and where the practical feasibility of substitution has not been tested.
- Opacity in fourth-party dependencies (the foundation models, training data sources, or sub-providers behind a vendor’s product) that may limit independent assessment of model behaviour, performance and resilience.

#### 4.3.3. Relevance for Actuarial Practice

Members may work across a mix of in-house models and systems provided by third-party vendors. This includes models embedded in third-party platforms and pre-trained foundation models. Members must evaluate these systems to determine whether they are fit for purpose, including testing the limitations of the system against risks driven by performance. Considerations span both the technical performance of the model and the broader technology ecosystem within which it sits.

This may represent a departure from traditional actuarial practice. Whereas Members may have historically built models, hosted them, and updated them on cycles they controlled, vendor AI systems are often updated on the vendor's schedule. Visibility of those update cycles may matter even without direct control over them.

#### 4.3.4. Questions a Member may consider asking

- What evidence of testing has the vendor provided? Has performance been independently validated?
- What contractual provisions are in place regarding system performance, model hosting, ownership of intellectual property, and training on customer data?
- What is the contingency plan if an AI system degrades or is discontinued? Are there sufficient resources to manage model migrations at short notice? What is the fallback plan if an AI system is unavailable due to infrastructure or platform failures?
- Where the organisation relies on a small number of AI vendors or foundation models across multiple critical use cases, has the feasibility of substitution or exit actually been tested, rather than only documented in principle?

### 4.4. Stakeholder management

#### 4.4.1. Context

AI systems do not sit in a vacuum. To be effective they need to be pointed at a real business problem, used by people who understand them, and accountable to the people they affect. Stakeholder management cuts across several AI6 themes and is often where AI deployments succeed or fail in practice. Poor stakeholder engagement at the start can mean solving the wrong problem; poor engagement at the end can mean affected customers have no meaningful route to challenge an outcome. This consideration is relevant across all stages of the Actuarial Control Cycle, with particular weight at problem definition and monitoring.

#### 4.4.2. Examples of Risk

- The operational risk of waste: for example, an AI system that is technically well-built but pointed at a problem the business did not actually have, because the people closest to the work were not consulted during design.

- The risk of overreliance: for example, users of an AI system who have not been trained on its limitations, leading them to over-rely on its outputs or fail to recognise when something has gone wrong.
- Compliance, reputational and / or customer experience risks due to disclosure, explainability, and contestability mechanisms that exist on paper but are not genuinely accessible to affected consumers. For example, appeal processes that require digital literacy the affected cohort does not have, or that are inadequately staffed to respond within reasonable timeframes.

#### 4.4.3. Relevance for Actuarial Practice

People are the critical ingredient to the successful rollout of an AI system. Members often sit at the intersection of technical teams, business stakeholders, and risk and governance functions, and may therefore be well placed to help organisations enact successful AI deployments.

Effective solution design starts with understanding the business problem, often by engaging stakeholders such as staff or customers. Regardless of whether the system is intended to augment knowledge work or improve customer experience, users need sufficient training to understand both how the system works and its limitations. Because no system is perfect, disclosure that AI is being used, explainability of outcomes where possible, and contestability mechanisms with appropriate redress are established concepts in AI governance that warrant practical attention rather than nominal compliance.

Members can act as translators between business, technical, and customer outcomes, smoothing knowledge gaps between these areas and providing advice on how to practically balance the trade-offs involved.

#### 4.4.4. Questions a Member may consider asking

- Who are the key stakeholders that should be engaged to understand the business context and problem that the AI system is intended to solve?
- What key assumptions are being made, and what is the impact if they are wrong?
- When something goes wrong, has the AI system been designed so that it is easy for a user to escalate to a human? Under what conditions is human-in-the-loop appropriate as a control to manage risks and escalations? What limitations do the humans-in-the-loop have, and how are those limitations managed?
- What disclosure, explainability, and contestability mechanisms are in place? Are they genuinely accessible to affected consumers, including those with language barriers or disabilities? Is there sufficient staffing to ensure consumer and user needs can be met?

# 5. Considerations of Professionalism

## 5.1. Competence and Care

Under the Code, a Member must have due regard to Professional Governance Material and Regulatory Guidance. While this Technical Paper is not compulsory, a Member may consider it while endeavouring to act with integrity, and to meet the competence and care requirements.

Under the Guidance to Support the Code of Conduct, the Member remains accountable for their work, regardless of the tools used to produce it. If an AI system produces an error that the Member relies upon, the accountability sits wholly with the Member; it cannot be outsourced to the AI system.

Accountability for AI systems may require Members to maintain technical competence. Professional Standard 1 (PS1) requires Members to maintain their knowledge and skills through continuing professional development; AI literacy may be an increasingly important part of that. A Member who cannot evaluate whether an AI model is fit for purpose may struggle to meet their existing professional responsibilities.

AI may also expand the capabilities of Members, enabling work that was previously impractical due to resource or skill limitations, for example, prototyping analytical pipelines or producing interfaces that visualise data or illustrate user experiences.

The ability to show rather than tell may strengthen a Member's capacity to influence stakeholders and bridge technical insight with commercial action. However, once a Member decides to avail themselves of new technology that unlocks new capabilities (such as creating visual interfaces), they are responsible for ensuring they have the right knowledge and skills in that domain to do so without compromising the quality of advice (e.g. to ensure there are no programming errors in the visual UI, which may historically have been outside of their domain or area of accountability).

Explainability is one area where the competence and care obligations may have particular bite. The appropriate level of explainability may depend on the use case. A targeting model for marketing may require relatively little; a pricing or underwriting model that drives consumer outcomes may require substantially more; an automatic process to decline service may require more again. Where a use case has been consciously assessed as not requiring full explainability, that may be a defensible professional position. Where explainability is likely to be required but is absent, a Member should consider whether a gap exists.

To act with care, a Member may take personal accountability for the elements of an AI system in which they have been involved. A Member may also consider encouraging the organisation to establish a suitably defined and accountable owner of the AI system as a whole, together with appropriate systems of governance and accountability.

Some organisations have begun to describe AI agents as if they were employees within the organisation. We reject this framing. AI agents are tools, and accountability must be clearly defined e.g. sitting with the human that deployed the agent, or some other accountable party. If use of an AI agent results in a breach, the organisation should be able to identify the accountable human role and the control failure that allowed it to occur.

### 5.2. Compliance and Speaking Up

The Code requires Members to comply with relevant laws, regulations and Professional Standards. There are many privacy, anti-discrimination, consumer protection and financial services laws that may be relevant to AI use, particularly where AI systems process personal or sensitive information, or make or influence decisions affecting individuals.

A Member has a responsibility to respond appropriately to non-compliance by others. This includes raising issues and concerns about aspects of an AI system as appropriate within their organisation. If concerns are not addressed within the organisation, a Member may need to consider what further escalation may be required.

### 5.3. Integrity and Objectivity

An important requirement of the objectivity principle of the Actuaries' Code is that a Member provides advice that is free from bias. If a Member is using AI to help form their advice, then, this may include considering biases that may exist in the historical data used to train a model, or biases that emerge through the design or deployment of the system.

It should be noted that some AI systems are not amenable to bias testing due to technical limitations, and even quantifying the amount and direction of bias in these systems is an emerging field of research. A Member may consider whether use of such systems is appropriate to each specific use case - with extra care taken in situations where strict regulatory obligations on the absence of bias exist.

A Member may also consider how to maintain independence to challenge AI outputs, particularly where the AI is built internally and the Member is employed by the same organisation, in the same way they would with other management-proposed assumptions but with potentially greater technical complexity.

A Member may use AI tools permitted by their organisation and follow internal policies on data confidentiality. Entering company or customer data into a public AI system may breach privacy obligations or contractual requirements, and a Member should consider whether such use is justified having regard to those obligations.

### 5.4. Communication and Documentation

Relevant considerations, including those discussed above, may be communicated and documented in line with a Member's responsibilities under the Code.

Many considerations in an AI context require careful judgement and trade-offs between different idealised outcomes. A Member may consider how such judgements and trade-offs are documented and communicated to appropriate individuals, particularly those in decision-making or responsible roles.

A Member should consider whether written documentation for an AI system is sufficiently comprehensive that a suitably qualified individual could understand the design, construction and operation of the system and offer comment or critique on its appropriateness.

AI systems may be complex and technical in nature, which can be particularly difficult for non-experts to understand. Members should take care to ensure that communication is tailored to the audience, and that technical detail does not obscure salient facts from being understood.

### 5.5. Further Guidance

The Institute is planning to release a publication on AI Risk Management in the Financial Services Sector, in the second half of 2026. Members may find this resource helpful to guide the operationalisation and practical application of AI Risk Management within their organisations. A link to this resource will be added when released.

## Appendix A: Illustrative Case Studies

The case studies below are illustrative scenarios mapping AI deployments to the AI6 themes and to the good practices set out in this Technical Paper. They are not real organisations, and they are not exhaustive. Their purpose is to illustrate how the practices set out above may apply in particular contexts. They draw primarily from insurance because that is where actuarial AI use is currently most mature, but analogous considerations likely apply elsewhere. A non-insurance example is included at the end to make that point explicit.

Each case study follows a consistent structure: scenario, description, the AI6 themes most directly engaged, the Member's role, key questions, and watch-outs.

## Case Study 1 AI-driven underwriting In Life Insurance

### Scenario

A life insurer implements an AI system that ingests application data, medical history, and third-party data to generate risk assessments. Standard applications receive automated decisions. Complex cases are escalated to human underwriters with the AI's recommendation attached.

### Description

AI underwriting may learn patterns from historical decisions, identifying risk factors that explicit rules may miss. The value is clear: faster decisions, lower acquisition costs, more consistent risk selection. But if past underwriting reflected biases, the AI may encode those biases as features. A model trained on a decade of decisions where certain postcodes or health conditions received less favourable terms may reproduce those patterns at scale.

The escalation process matters as much as the model. When the AI provides a recommendation with a confidence score, human underwriters may anchor to it. A relevant question is whether the underwriter is genuinely exercising judgement or simply providing a veneer of oversight.

### Relevant AI6 themes

- **Measure and manage risks:** the risk profile depends on the AI's autonomy. Generating risk scores for human review is lower-risk. Automatically declining applications without review is higher-risk. While it may be the same model, there are different governance considerations.
- **Understand impacts:** the AI may systematically disadvantage people with mental health histories, genetic predispositions, or disabilities. Applicants may benefit from a clear pathway to challenge AI-influenced decisions and to receive a meaningful explanation.
- **Test and monitor:** monitoring decision distributions across demographic groups may be useful. If the automated decline rate shifts for a particular cohort, that may warrant investigation.

### The Member's role

If the AI shifts the risk profile of new business, the Member's assumptions about expected claims, lapse rates, and mortality or morbidity may need recalibration. The actuarial team may also be well-positioned to design bias-detection frameworks, as statistical testing across demographic groups is a familiar actuarial skill. The Actuaries' Code obligations on integrity and care may be engaged where a system systematically disadvantages a vulnerable group.

### Key questions

- Does the training data reflect underwriting decisions the insurer would endorse if reviewed today?
- What is the effective override rate for human underwriters, and is it tracked?
- How does the model handle applicants with characteristics not well-represented in training data?

### Watch-outs

- **Encoded bias.** The model reproduces historical biases at scale and speed.
- **Proxy discrimination.** Individually neutral variables (postcode, occupation) collectively approximate protected characteristics.
- **Nominal oversight.** Underwriters anchor to the AI recommendation; the override mechanism exists but is rarely used.

AI-assisted underwriting is in use across the Australian market, though governance maturity varies. The key gap is not the AI capability but the monitoring and bias-testing infrastructure around it.

## Case study 2: AI-automated claims processing

### Scenario

An insurer deploys AI to streamline claims processing, automatically approving straightforward claims and flagging complex or unusual ones for human review. The system assesses documentation, cross-references policy terms, checks for fraud indicators, and generates a recommended action with a confidence score.

### Description

Claims automation may deliver immediate value: faster processing, lower costs, more consistent application of terms. But an incorrect automated denial may have a different character from an incorrect approval: the insurer can recover an overpayment, but a wrongly denied claimant may face financial hardship while pursuing review.

The fraud detection dimension adds complexity. AI may identify genuine fraud patterns, but false positives carry specific harm: the claimant is treated as a suspected fraudster. If fraud indicators correlate with ethnicity, language, or geography, the system may impose disproportionate scrutiny on certain communities.

### Relevant AI6 themes

- **Maintain human control:** claims assessors reviewing escalated cases may need authority, information, and time to exercise genuine judgement. If KPIs incentivise throughput,

oversight may become nominal. Contingency planning matters too: if the system fails during a catastrophe event, can claims continue manually?

- **Test and monitor:** approval rates, denial rates, and appeal outcomes across claimant demographics may warrant monitoring. If claimants from certain cohorts are disproportionately flagged, that pattern may warrant investigation of the model.
- **Share essential information:** a claimant whose claim is denied may benefit from understanding whether AI influenced the decision and what they can do about it.

### The Member's role

Claims data feeds directly into reserving, pricing, and capital models. If AI changes claim decision patterns, downstream actuarial assumptions are potentially affected. The Appointed Actuary should consider how AI is changing claims distributions, and Members may consider seeking arrangements where AI-influenced claims are flagged in the data warehouse. Without this, experience analysis may struggle to distinguish changes in underlying risk from changes in claims handling.

### Key questions

- What is the false denial rate, and does it vary across claimant cohorts?
- How are AI-influenced claims flagged in data so the actuarial team can isolate effects on experience analysis?
- What is the contingency plan during a catastrophe event when claims volumes spike?

### Watch-outs

- **Asymmetric harm.** The system may be optimised for average accuracy while producing unacceptable false denial rates for edge cases.
- **Opacity to the actuarial team.** Claims distributions change but the actuarial team cannot access the AI decision logic or identify AI-influenced claims in their data.
- **Complaint escalation.** Automated denials without clear explanation may generate complaints that escalate to AFCA, creating cost and reputational damage disproportionate to individual claim value.

## Case study 3: AI pricing models across the risk spectrum

### Scenario

An insurer develops an AI pricing model for motor insurance deployed across three contexts: generating indicative quotes on a comparison website (lower-risk), setting actual premiums with human oversight (medium-risk), and automatically declining high-risk applications without review (higher-risk).

### Description

AI pricing may augment traditional GLMs by identifying non-linear relationships and micro-segments, enabling better risk differentiation. But the same capability that improves accuracy may also enable hyper-segmentation that affects the pooling function of insurance. If the AI prices a cluster of potentially high-cost customers out of the market, the social function of insurance may be eroded.

A key governance consideration is that context may determine risk, not the model itself. An error in an indicative quote may be minor. An error in a binding premium affects cost. An automatic decline excludes someone from coverage entirely. The same model can have three different governance considerations.

### Relevant AI6 themes

- **Measure and manage risks:** the risk level depends on deployment context, not the model itself. Governance may be proportionate: indicative quotes may warrant lighter oversight than automatic declines. This context-dependent assessment is central to AI6 and aligns naturally with actuarial model risk management.
- **Understand impacts:** a customer receiving a premium increase may benefit from understanding whether AI influenced it and what factors contributed. Customers priced out of coverage or automatically declined may experience the most severe impacts.
- **Decide who is accountable:** the pricing actuary may sign off on the rating structure, but the AI may operate within it in ways that are not fully explainable. When a decision is challenged, the accountability chain may need to be clear.

### The Member's role

The Member may consider whether the AI produces outcomes consistent with actuarial principles. This may include understanding not just aggregate performance but behaviour at the margins: edge cases, segment-level fairness, and whether outcomes are defensible under anti-discrimination law. An AI model that optimises profitability by excluding the highest-risk customers may be commercially rational but professionally questionable if it undermines insurance's social function.

### Key questions

- Can the Member explain the key premium drivers for any individual customer?
- At what premium level does the model's output cross from risk selection into effective exclusion from coverage?
- How is model drift detected, and what triggers a review versus routine recalibration?

### Watch-outs

- **Hyper-segmentation eroding the pool.** AI identifies micro-segments and prices them out, affecting risk-pooling.
- **Proxy discrimination.** Postcode, vehicle type, and credit data may together approximate ethnicity or socio-economic status.
- **Explainability gap.** The Member certifies the pricing basis but cannot explain the premium for a specific customer.

### Case study 4: AI-predicted claim costs for reserving

### Scenario

An insurer implements AI to predict claim costs for reserving. The system analyses historical claims data, identifying development patterns that traditional chain-ladder and Bornhuetter-Ferguson methods may not capture. It generates estimated ultimate costs at individual claim level, aggregated to produce portfolio-level reserves. The Appointed Actuary uses AI estimates alongside traditional methods in the valuation.

### Description

The value may be strongest for long-tail classes where traditional methods struggle, such as liability lines with volatile development, or workers' compensation with complex return-to-work dynamics. If the AI identifies early indicators of adversely developing claims, the actuarial team may establish more accurate reserves earlier.

But the feedback loop may be slow. In both reserving and pricing, the true accuracy of AI predictions may not be known for years. Reliance is placed on a model whose performance cannot be fully validated within normal reporting cycles.

A specific governance consideration is that the AI may produce estimates that management finds convenient. Lower reserves than traditional methods may invite preference for the AI; higher reserves may invite preference for the traditional approach. Members should be alert to selective reliance.

### Relevant AI6 themes

- **Maintain human control:** the Appointed Actuary retains professional authority to adjust, override, or disregard AI estimates. The valuation process need not make the AI's output the default with the actuarial role reduced to approving at the margin.
- **Test and monitor:** validating AI estimates against actual development for historical cohorts, and tracking prediction accuracy by claim type, development period, and severity

band, may be useful. Unlike pricing models refreshed annually, reserving models may be relied upon continuously, so drift detection may need to be correspondingly continuous.

### The Member's role

The Appointed Actuary's valuation opinion is a cornerstone of insurance regulation. If AI contributes to the estimate, understanding the model's methodology, data, limitations, and the circumstances under which it performs poorly are essential. Sensitivity analysis may help: how much does the aggregate reserve change if AI estimates are adjusted by plausible error margins?

Where the AI is built internally and the Appointed Actuary is employed by the same insurer, the Member may need to maintain independence to challenge the AI's outputs, in the same way they would with management-proposed assumptions, but with potentially greater technical complexity.

### Key questions

- How do AI estimates compare to traditional methods across claim types, development years, and severity bands?
- Can the Member explain why the AI differs from traditional methods for any material segment?
- Is there a risk that management selectively adopts the AI estimate when it produces a favourable result?

### Watch-outs

- **Selective reliance.** The AI estimate is adopted when it supports lower reserves and disregarded when it suggests higher ones.
- **False precision.** Individual claim estimates may look precise but obscure the true uncertainty in the portfolio-level reserve.
- **Independence pressure.** Technical complexity may make it harder for the Appointed Actuary to challenge AI outputs, particularly where AI expertise within the actuarial team is less developed than within the data science team.

## Case study 5: AI-influenced premium communications

### Scenario

A customer receives a significant home insurance premium increase at renewal. The pricing model incorporates AI-driven risk assessment using climate modelling, property-level data, and claims history. The customer asks why. The service representative cannot explain the specific factors because the AI's contribution has not been translated into language that a layperson can easily understand. (Such a scenario may play out where tools like GLMs are already used).

### Description

This may test transparency in its most customer-facing form. While insurers may be willing to explain changes to customers, the AI's contribution may be difficult to translate. Traditional rating factors (age, location, sum insured) are intuitive. AI-derived adjustments based on property-level hazard modelling or neighbourhood claims patterns may be less so.

Transparency may be a design requirement, rather than a communications problem to be solved downstream. Where the model cannot explain its output in terms a customer could reasonably understand, the governance framework may be incomplete.

### Relevant AI6 themes

- **Share essential information:** the central theme here. Customers may benefit from understanding whether AI influenced their premium, what factors contributed, and what they can do about it. Customer-facing staff may need access to plain-language explanations.
- **Maintain human control:** customers may benefit from being able to request human review by someone with sufficient understanding of the model to assess reasonableness, as opposed to only confirming the system produced the correct output.

### The Member's role

Members involved in the design of the pricing model may sit as the natural bridge between the AI model and the customer. Where the Member cannot explain why the model produced a specific premium, this may signal insufficient understanding of the model. A Member may consider advocating for explainability as a design requirement, even where this involves trading off some predictive power. That trade-off is a professional judgement call Members may be well-positioned to make.

### Key questions

- Can the insurer explain, in plain language, the key factors driving any individual customer's premium?
- What information do customer service representatives have about the AI's contribution?
- If the model produces a premium the Member would not have expected, does the Member have information and authority to investigate?

### Watch-outs

- **Explanation gap.** Premiums the insurer cannot explain may create complaints and reputational damage disproportionate to the amount involved.

- **Availability impacts.** AI pricing may systematically increase premiums for high-risk properties until insurance becomes unaffordable for entire communities creating a social and political consideration as opposed to solving only a commercial one.

### Case study 6: AI-recommended claim decline on policy wording

#### Scenario

An AI system recommends declining a business interruption claim based on its interpretation of policy wording. The system is trained on historical claims, policy documents, and past adjudication outcomes. The claim assessor receives the recommendation alongside cited policy clauses and historical precedents. They must decide whether to accept or override.

#### Description

Policy wording interpretation may be inherently contextual. The same clause may apply differently depending on circumstances, the insured's reasonable expectations, and case law. AI trained on historical outcomes may identify patterns but may not exercise the contextual judgement complex claims require.

The assessor receives an authoritative-looking recommendation supported by cited clauses. This framing may be difficult to override, particularly under time pressure. Even where the AI is highly accurate on average, the proportion of claimants wrongly declined may face insolvency, unrepaired property, or lost livelihoods. Individual harm is not necessarily mitigated by aggregate accuracy. The COVID-19 business interruption test cases (2020-2022) showed that even experienced professionals can disagree on wording interpretation; AI trained on pre-pandemic outcomes would likely have been systematically wrong on these claims.

#### Relevant AI6 themes

- **Maintain human control** The assessor may need authority, information, and training to override. However, this will require understanding of the claim's specifics, the AI's limitations, and clear authority to override without facing pressure to accept. If the assessor is not given enough time, oversight may be nominal.
- **Decide who is accountable:** when a declined claim is overturned on review or by AFCA, accountability may need to be traceable across the assessor, claims manager, and data science team.

#### The Member's role

If AI-recommended declines increase the denial rate for a specific class, reserving assumptions may need adjustment. Where those declines are subsequently overturned, the Member may need to factor in reversal rates and dispute resolution costs. The Member's perspective may be valuable

in testing whether AI recommendations are consistent with the insurer's claims philosophy and regulatory obligations.

### Key questions

- What is the override rate for AI-recommended declines, and is it analysed?
- How do appeal and AFCA outcomes for AI-recommended declines compare to human-only decisions?
- How does the AI handle novel scenarios materially different from training data?

### Watch-outs

- **Anchoring.** The AI's recommendation, supported by cited clauses, creates a frame assessors may be reluctant to challenge.
- **Novel scenarios.** AI may not anticipate loss types outside its training data. When these emerge, recommendations may be systematically wrong.
- **Reputational cascade.** Systematic AI-recommended declines overturned on appeal may attract regulatory scrutiny and media attention exceeding the financial impact of individual claims.

## Case study 7: AI-driven marketing and customer targeting (non-insurance)

### Scenario

A retailer deploys an AI system to personalise offers and pricing for customers in real time, drawing on internet browsing behaviour, purchase history, and demographic data. The system serves different prices and promotions to different customers.

### Description

Marketing and targeting use cases may sit at the lower-stakes end of the AI spectrum: an inappropriate offer for a particular customer is rarely catastrophic. But the same governance considerations may still apply, and at scale, marketing AI may quietly produce outcomes that would be unacceptable if surfaced. Personalised pricing that systematically charges more to customers in particular postcodes, or offers that exclude particular cohorts from promotions, may attract regulatory and reputational scrutiny that the underlying business case did not anticipate.

The lower stakes mean the bar for explainability and human oversight may be lower than in pricing or claims, but it is not zero. The governance consideration may be proportionality, not absence.

### Relevant AI6 themes

- **Measure and manage risks:** low stakes do not mean no stakes. Risk screening may still differentiate between, say, a recommendation engine and dynamic price personalisation, which raise different regulatory and reputational considerations.
- **Understand impacts:** targeting and personalisation may produce systematic differences in offers across demographic groups even where individual decisions look benign.

### Why this example is included

AI6 applies wherever AI is deployed, not only in high-stakes actuarial contexts. The good practices set out in this Technical Paper translate to customer-facing personalisation, dynamic pricing in retail, supply chain optimisation, and other domains in which Members are increasingly working. The adjustment is one of calibration: the depth of governance may be proportionate to the stakes, while the underlying considerations do not change.

## END OF TECHNICAL PAPER

### Document control box

| Version | Title of Document | Name of approving Council or Committee | Date of approval | Date of publication |
|---|---|---|---|---|
| 1.0 | Technical Paper: Considerations for AI adoption - Applying the AI6 | Professional Practice Committee | 30 July 2026 | 21 August 2026 |