Perspectives on Cyber Risk 2026

The AI edition

MinterEllison.

Contents

Foreword

This is the 11th edition of MinterEllison's *Perspectives on Cyber Risk*

Paul Kallenbach
Partner
National Legal Cyber Leader

2026 Australian Cyber Lawyer of the Year

Each year, for over a decade, we have asked Australian senior decision-makers how they see cyber risk, what they are doing about it, and from where they think the next test will come. The themes change each year. However, the underlying question is the same: how do we run a digitally-enabled business when those trying to compromise it are getting faster and more capable?

This year, the answer is shaped by one factor above all others: artificial intelligence (**AI**). The defining feature of the 2026 cyber environment is the speed at which AI is reshaping both the threat landscape and the defence toolkit. AI has reached near-universal adoption across the cohort we surveyed, with 98% introducing AI in the last 24 months. Threat actors are equally AI-enabled. AI-enabled threats now sit alongside ransomware as the leading concern identified by our respondents; three of the top five named concerns are AI-adjacent. The threat surface is not just larger, it is mutating, as AI accelerates concentrated attacks, exposes AI systems themselves as targets, and increasingly acts as the attacker with limited human involvement.

While AI has featured on both the attack and defence sides of cyber risk for years, what has changed in 2026 is the capability of frontier models and the scale that capability enables. In its April 2026 Letter to Industry, the Australian Prudential Regulation Authority (**APRA**) highlighted the cyber threat posed by high-capability frontier AI models, directly referenced Anthropic's Mythos, and called for a 'step change' in how organisations manage AI-related risk. The Australian Securities and Investments Commission (**ASIC**) echoed the warning weeks later, stating that frontier AI was '*intensifying the cyber risk environment*'. Most recently, the Five Eyes cyber security agencies issued a rare joint statement calling on leaders to act urgently as AI transforms cyber risk. The regulatory message is clear: AI is changing the speed and scale at which vulnerabilities can be found and exploited and organisations must take steps to address this.

Against this backdrop, cyber incidents are now firmly a 'when' not 'if' reality. Our 2026 data leaves little room for doubt: 71% of the organisations we surveyed experienced at least one cyber incident in the last 12 months. No level of preparedness will ever fully inoculate an organisation against compromise.

However, preparedness does enable organisations to:

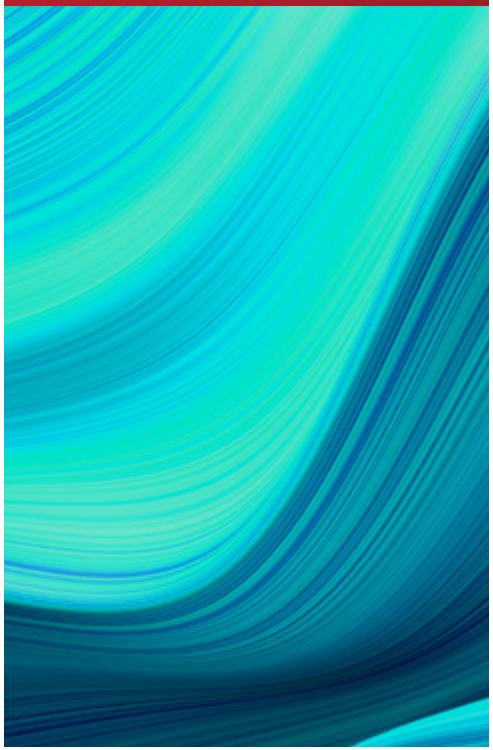
- successfully navigate what happens after a breach occurs, operationally, legally and culturally;
- act quickly to mitigate the impact, cost and loss arising from an incident; and
- ensure their actions are defensible.

The regulatory landscape has also strengthened materially in the last 12 months. New obligations spanning privacy, cyber security, operational resilience and AI governance have increased the legal exposure that follows a poorly managed incident. Regulators are moving from supervision to enforcement, and the organisations best positioned are those that can demonstrate defensible, documented AI and cyber governance. We discuss these regulatory developments in detail later in this report.

We hope this report assists directors, general counsel and CISOs to think about where their attention is best focused in the year ahead, and crystallises the key issues that warrant immediate action.

Executive summary

AI, cyber risk and the accountability gap



Cyber risk has evolved, changing the way attacks are conceived and scaled, compressing attack timelines, and straining traditional control models. AI is the variable that has accelerated that shift most visibly over the last 12 months, and the legal and regulatory framework has hardened alongside it.

The result is an accountability gap between the pace of the threat and the governance structures on which most organisations rely.

In FY2024-25, the Australian Signals Directorate's (**ASD**) Australian Cyber Security Centre (**ACSC**) received more than 84,700 cybercrime reports on average, one every six minutes. It responded to over 1,200 cyber security incidents, an 11% increase year on year, and issued more than 1,700 proactive notifications about malicious cyber activity, up 83%. The average cost of cybercrime to large Australian businesses rose to $202,700 an incident, up 219% from the previous year. Denial-of-service activity rose by more than 280%, and critical infrastructure now accounts for 13% of all reported incidents.[1]

Our 11th annual cyber survey looks beneath those headline figures. Across 150 senior decision-makers from Australian-headquartered organisations, 71% have experienced a cyber incident in the past 12 months, 64% through a direct attack and 57% through a supplier or vendor breach.

What has shifted most over the last 12 months is the threat surface itself. AI-enabled threats are now named the leading cyber concern by 25% of survey respondents, second only to ransomware at 30%. This is the first time in this survey series that AI has challenged ransomware for the top position. Phishing, the other top-five concern, is increasingly AI-augmented. AI adoption has kept pace: 98% of respondents have introduced AI in the past 24 months and 95% report a formal AI governance framework. AI has moved from emerging technology to default enterprise infrastructure inside a single survey cycle.

> *"Organisations should operate with a cyber mindset, one that assumes compromise and prioritises response. The defining question is no longer whether an incident will occur, but whether the response is built for the way incidents now occur. With the widespread adoption of AI, cyber readiness must extend to comprehensive AI governance."*

Paul Kallenbach
Partner, Technology, Digital & Data, National Legal Cyber Leader

"Emerging AI threats are challenging even the most sophisticated organisations. More needs to be done to uplift enterprise-wide AI governance to manage these threats and meet regulator expectations."

Chelsea Gordon
Legal Lead – AI Advisory

Respondents present a confident posture; 98% have benchmarked their cyber maturity against an established framework, 91% have an incident response plan and 51% test at least quarterly. The harder question is whether organisations' plans have kept pace with the way attacks now happen, including deepfake-enabled fraud, prompt-injection vulnerabilities, supplier-platform compromise and agentic offensive tooling. Our survey responses suggest that in many cases they have not.

When asked to describe AI cyber risk in a single word, 32% of respondents reached for anxiety-laden language (worried, anxious, alarmed, overwhelmed, vulnerable), while only around 8% chose net-positive words. The confidence organisations report in their response frameworks sits uneasily beside the discomfort they express about the threats those frameworks now face.

The legal and regulatory environment has hardened in step. Mandatory ransomware payment reporting came into force on 30 May 2025. The statutory tort for serious invasions of privacy commenced on 10 June 2025, allowing individuals to sue for intrusion upon seclusion or misuse of personal information, and materially increasing post-incident litigation risk.[2] APRA's CPS 230 commenced on 1 July 2025, raising the bar on operational risk and the management of material service providers.[3]

In its December 2025 National AI Plan, the Australian Government confirmed it would rely on existing technology-neutral laws rather than introduce AI-specific mandatory guardrails. That position has since shifted: on 15 July 2026, the Prime Minister announced mandatory Australian Standards for AI, a new Office of AI within the Department of the Prime Minister and Cabinet, and legislation flagged for early 2027, though the mandatory obligations foreshadowed so far are limited to large data centres and copyright protection rather than an economy-wide regime. Pending that detail, the practical task for organisations is unchanged: map and operationalise the duties they already owe under existing law, while also tracking the design of the new mandatory standards over the coming year.[4] What is already locked in is that, from 10 December 2026, a new transparency obligation will require APP entities to disclose in their privacy policies where they use personal information in automated systems to make, or substantially assist in making, decisions that could significantly affect an individual's rights or interests.[5]

Regulators are not waiting for new legislation. In its 30 April 2026 letter to regulated entities, APRA observed that *'assurance practices are not keeping pace with the scale, speed and complexity of AI'*.[6] The Office of the Australian Information Commissioner (**OAIC**) has intensified its enforcement activity, including in relation to AI-enabled technologies such as facial recognition.[7] In its cyber letter of 8 May 2026, ASIC called on regulated entities to strengthen their cyber resilience as frontier AI intensifies the global cyber risk environment.

Taken together, these reforms (and the stance of the regulators) increase the accountability gap. The legal exposure that follows a poorly managed incident is materially greater than it was 12 months ago.

This year's edition of *Perspectives on Cyber Risk* is structured around three themes drawn from our survey findings, and the current and evolving regulatory landscape, each with practical priorities for Boards and leaders.

- **Theme 1:** Incidents have become the norm, and the threat landscape is shifting.
- **Theme 2**: AI is enterprise infrastructure, and governance must catch up.
- **Theme 3:** Organisations are preparing for the wrong incident.

In 2026, cyber preparedness means less about whether an incident occurs and more about whether the response holds up to scrutiny after the fact.

With threats now moving at machine speed and the regulatory environment more exacting than it was a year ago, resilience is not just operational, but a Board-level responsibility.

At a glance: the 2026 picture

Cyber on the leadership agenda

48% rank cyber in their **top 5** corporate risks; a further **39%** rank it **medium**

Top two named cyber concerns

- Ransomware (**30%**)
- AI-enabled threats (**25%**)

The incident surface

71% experienced at least one cyber incident in the last 12 months

57% reported a cyber incident affecting a third-party supplier or vendor vs **50%** in 2025.

Preparedness in practice

91% have an incident response plan and regularly test or rehearse that plan.

77% conduct cyber awareness training frequently, up from **70%** in 2025. (**2026: 26%** monthly, **51%** quarterly. Only **7%** test annually or less).

98% have measured their cyber security maturity against an established framework.

AI: now baseline

98% have adopted AI in the past 24 months

up from **70%** adopted AI in 2025.

95% have a formal AI governance framework.

41% rank cyber security as their number-one AI risk, ahead of privacy and data protection (**23%**), operational and model failure (**17%**) and legal and regulatory risk (**14%**).

64% are very confident their organisation has appropriate controls to manage AI-related risks.

Confidence in control

98% believe they are sufficiently staffed to monitor and manage cyber security.

67% are very confident they know what data they hold, where it resides, and who can access it.

65% are very confident in understanding regulatory, contractual and legal exposure.

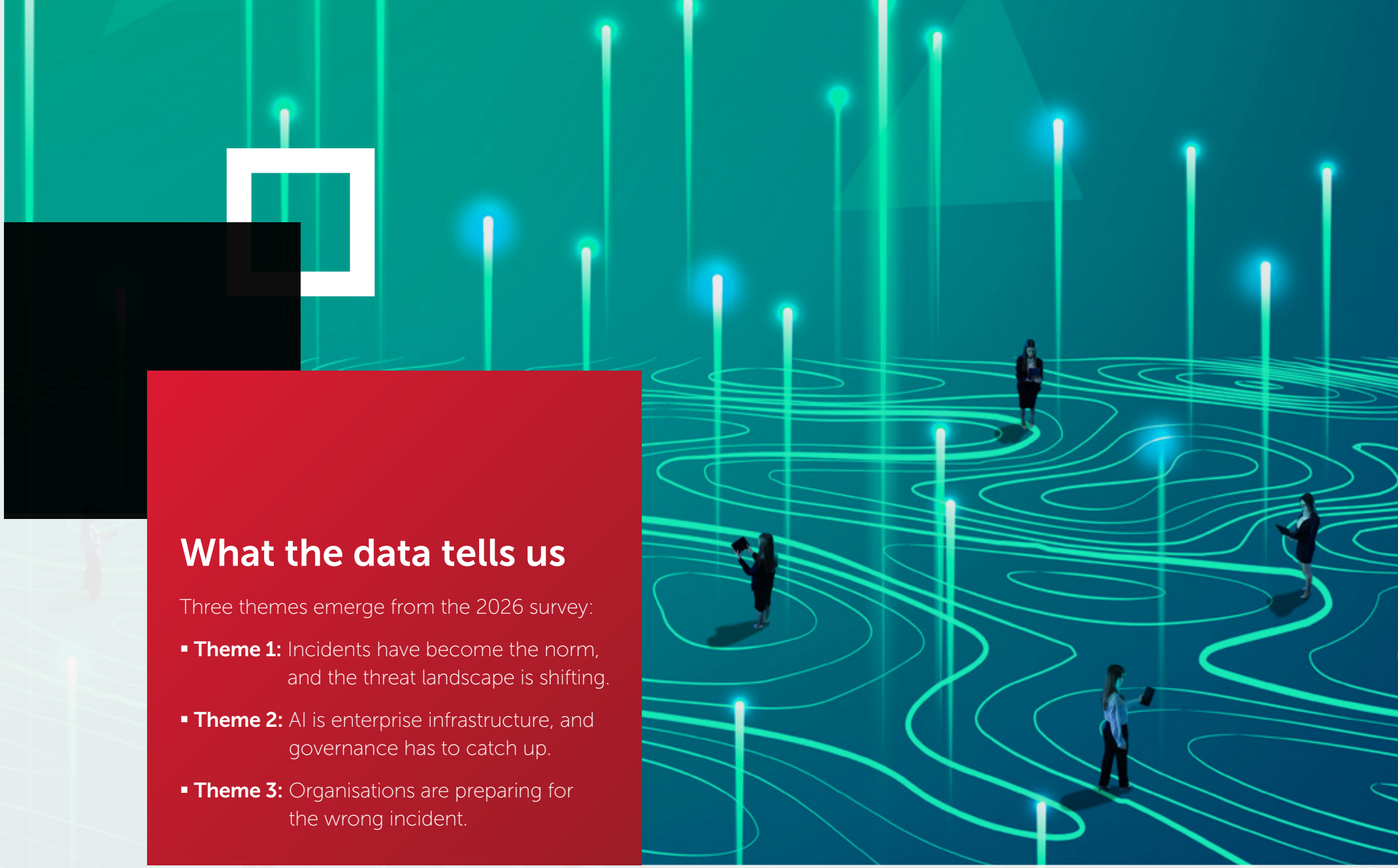
94% hold cyber insurance.

Sentiment

32% described AI-cyber risk using anxiety-laden language; ~15% used cautious/wary language; ~8% used net-positive language.

In one word, how does your organisation feel right now about the pace of AI-driven cyber risk?

- Concerned
- Cautious
- Alarmed
- Overwhelmed
- Anxious
- Vigilant
- Prepared
- Confident
- Rapidly evolving
- Proactive



What the data tells us

Three themes emerge from the 2026 survey:

- **Theme 1:** Incidents have become the norm, and the threat landscape is shifting.
- **Theme 2:** AI is enterprise infrastructure, and governance has to catch up.
- **Theme 3:** Organisations are preparing for the wrong incident.

What the data tells us

Theme 1:

Incidents have become the norm, and the threat landscape is shifting

What has shifted in the last 12 months is not the fact of incidents, but the forces driving them, AI has decoupled the scale of an attack from the effort required to mount it, and our survey respondents recognise this.

The productivity and automation gains for organisations are real, but so is the uplift in capability it hands to threat actors.

In this report, we explore three amplified roles that AI plays in the cyber security landscape, fuelled by the increasing capabilities of frontier models.

First, new capabilities have made conventional attacks like phishing and deepfake fraud faster and more convincing. Second, the AI systems organisations are deploying have themselves become targets, exposed to prompt injection, data poisoning and model manipulation. Third, and most recently, AI has begun to act as the attacker, capable of running an intrusion with little human direction.

Cyber incidents themselves are now a routine operating reality.

Across our respondent base, **71%** experienced a cyber incident in the past 12 months, **64%** through a direct attack and **57%** through a supplier or vendor breach.

Our survey data illustrates just how quickly AI has reshaped the threat hierarchy. For the first time in this series, AI-enabled threats are challenging ransomware as the leading cyber concern: 25% of organisations name them as their primary concern, second only to ransomware at 30% and ahead of every other named threat. (Ransomware itself is being accelerated by AI-enabled threats, from AI-crafted initial access through to faster exploitation and more convincing extortion.)

Cyber security – including attacks on and manipulation of AI systems – is now ranked the number one AI risk by 41% of respondents, ahead of privacy at 23%, which our 2025 edition reported as the leading AI concern. But the shift is not only in what attackers can do – it is also in what organisations have exposed. With 98% of organisations having adopted AI in the past 24 months, the attack surface has expanded correspondingly: more endpoints, more data flows, more systems whose behaviour can be changed without a line of code being touched.

Alongside rapid AI adoption has been the step-change development in capability of frontier AI models. In its 30 April 2026 letter, APRA confirmed it is engaging across the sector on the potential for increased cyber threats from high-capability frontier models, naming Anthropic's Mythos directly, and APRA observed '*clear recognition from regulated entities of the need for a step change in cyber practices*'.[8]

ASIC echoed this sentiment in its 8 May 2026 letter and accompanying media release, warning that '*frontier artificial intelligence (AI) intensifies the global cyber risk environment*'.[9] The capabilities that once required a skilled human team - reconnaissance, exploit development and social engineering at scale - are now within reach of, and are increasingly within the capability of, an AI model acting on its own.

That expanded attack surface has structural implications for every organisation in our survey – including the 91% of respondents that maintain and regularly test an incident response plan. Most defensive controls, governance models and assurance processes have been designed for threats that evolve incrementally and at human speed, and were tested at fixed points in time.

Frontier AI models have dramatically compressed the time attackers need to go from discovering a vulnerability to exploiting it, and have driven the marginal cost of each attempt close to zero, so a bad actor needs only one exploitable seam. The result is a widening gap between offence and defence that existing assurance models, built on slower assumptions, were not designed to address.

Two patterns in particular stand out.

PATTERN 1 › **The supply chain has become the single largest source of breach exposure.**

In our recent survey, more than half **(57%)** of organisations reported a cyber incident that occurred due to an attack on a third-party supplier or vendor in the past 12 months – **up from 50%** in our 2025 edition.

Once cumulative exposure is accounted for, with the same platform breach rippling across dozens of downstream customers, the true impact is considerably larger than the direct-attack rate suggests.

The Snowflake campaign illustrates this pattern. From April 2024, attackers used stolen credentials to access customer tenancies on the Snowflake data platform, exploiting the absence of enforced multi-factor authentication. The breach cascaded through that year and into 2025, surfacing at Ticketmaster, Santander, AT&T and others. Nothing about the underlying cause was unusual: a widely used platform, reused credentials and a missing control. But the downstream effect was significant, hundreds of millions of records exposed, and detection across affected organisations patchy and slow.

The lesson is that an organisation's real cyber risk exposure is not solely a product of its own controls, but of the controls it requires across its supply chain. Regulators now expect organisations to act on that lesson. APRA has stated that it expects regulated entities to take proactive steps to manage supplier risks, which must include, at minimum:

- mapping and maintaining visibility over the full AI supply chain;
- contractual and governance arrangements;
- a clear understanding of model behaviour, behavioural changes, performance issues and outcomes; and
- active management of concentration risk.

For APRA-regulated entities, these expectations are underpinned by CPS 230 (which came into force on 1 July 2025), requiring those entities to identify their material suppliers, document the risks those suppliers introduce, and demonstrate that they have taken appropriate measures to manage those risks, and that their continuity plans work.

ASIC has also urgently reminded its regulated entities to uplift their practices to actively manage third-party risks, including where services introduce concentration of systemic exposure.

Similarly, from April 2025, the *Security of Critical Infrastructure Act 2018* (Cth) (**SOCI Act**) requires regulated entities to manage data storage systems as critical infrastructure – which necessarily requires proactive management of third-party (including hyperscaler) related risks.

The underlying expectations here are substantive: ASIC, APRA and SOCI Act-regulated organisations must be able to explain, in writing and on request, how they identify supplier risk, how they classify risk, what their contracts require, and how they monitor risk over time. That raises a higher bar for what has historically been expected of most third-party risk programs, and applies whether or not a regulator comes knocking.

This is where risk management, AI governance and cyber risk management converge. MinterEllison is increasingly supporting clients to build a single, comprehensive governance program that manages third-party AI and cyber risk consistently across their:

- organisational risk management framework;
- procurement and third-party supply-chain risk programs;
- AI system register; and
- SOCI Act-mandated Critical Infrastructure Risk Management Program,

with oversight flowing through executive management to the Board. This integrated approach enables organisations to demonstrate the defensible, documented governance that regulators now expect.

PATTERN 2 › **AI-assisted attacks have gone mainstream**

The UK's National Cyber Security Centre captures it well: AI is not reinventing cyber-attacks, but it is making them faster, cheaper and easier to run at scale.[10] The impact on organisations manifests in three ways:

Compressed timelines

- AI compresses the loop between vulnerability discovery, weaponisation and exploitation. What once took attackers days or weeks can now be completed in minutes: leaked source code analysed, weaknesses identified, exploit code or convincing phishing content generated at scale. Patching cycles, manual review queues and quarterly governance rhythms were built for a slower environment. In social engineering especially, the time-to-exploit is now measured in minutes or hours, not weeks.

Reduced barriers to entry

- AI removes the coordination limits that once constrained large-scale attacks. Activities that previously required teams of skilled operators can now be executed by a single person using AI-enabled tooling – automating reconnaissance across thousands of targets, profiling systems and vulnerabilities, and launching tailored attacks against entire sectors simultaneously. The scaling occurs without a corresponding increase in cost or effort: the number of victims grows, but the attacker's workload does not. That shift disproportionately affects governments and large enterprises whose broad digital footprints and extensive user bases make them efficient targets for mass exploitation.

Higher first-attempt success rates

- AI does not just automate attacks, but optimises them. Personalisation and contextual plausibility improve across phishing, voice-cloned scams, business email compromise and deepfake fraud. The 2024 Arup incident remains the starkest example: a finance employee transferred USD 25 million after a video call with what appeared to be the company's CFO and several colleagues – all of which turned out to be AI-generated. Attacks like these require no sophisticated infrastructure – just a competent operator, readily available technology, a convincing scenario, and a target focused on the wrong signals.[11]

A further striking demonstration of where these capabilities now stand stemmed from a research disclosure rather than a criminal gang. In late February 2026, security startup CodeWall turned an autonomous offensive AI agent loose on McKinsey's internal AI platform, Lilli – a tool used by over 70% of the firm's roughly 43,000 employees.[12] The agent had no credentials, no insider knowledge, and no human directing it in real time.

Within two hours, the agent had discovered 22 unauthenticated Application Programming Interface (**API**) endpoints, exploited a Structured Query Language (**SQL**) injection vulnerability in one of them, and gained full read-and-write access to a production database containing 46.5 million internal chat messages, 728,000 client files, and 57,000 user accounts.

The data exposure alone would have been serious. The write access was the more dangerous problem. The agent could reach Lilli's 95 system prompts – the instructions that govern how the chatbot responds to consultants' queries. An attacker could have rewritten those instructions across the entire platform, without deploying code or triggering a deployment alert, just surreptitiously changing the answers on which tens of thousands of employees relied.

The risk here is no longer confined to data leaving the building. It extends to an attacker rewriting the tools that people trust quietly across many systems at once, and without tripping a single alert.

Here, one autonomous agent accomplished all of that, start to finish, with no human in the loop.

The Lilli disclosure demonstrates what is operationally possible. Two observations from it matters for Boards:

- AI is not inventing vulnerabilities, but is rapidly finding and exploiting them. The vulnerability class in this case – an SQL injection – was more than twenty years old. An autonomous agent found and exploited it in under two hours, on a platform that had been live for two years, with those end points undetected.
- Internal AI platforms carry a particular risk. They tend to hold large volumes of sensitive material, strategic discussions, client work, proprietary research, and their behaviour is governed by configuration (system prompts, retrieval indexes) that sits within the same infrastructure. Once compromised, these platforms have not just suffered a data breach, but an integrity breach, since the answers the platform gives can no longer be trusted.

The Lilli incident distils the three patterns explored in this theme:

A supply chain attack surface, an AI compressed timeline, and the quiet manipulation of a trusted internal tool. For Boards, the lesson is that governance, assurance and response need to keep pace with a threat landscape that has already moved.

The starting point is with how an organisation governs the AI systems it has already deployed. As Theme 2 explores, 95% of survey respondents report having a formal AI governance framework, but having a framework on paper is not the same as governance that has already been operationalised and can withstand regulatory scrutiny.

"The speed and scale of possible cyber threat vectors have changed fundamentally in the last 12 months. While the key cyber risks facing organisations are not necessarily new, the marginal cost of each attempt is close to zero – such that a single actor can launch hundreds or thousands of attacks at once. This creates a leadership imperative to act."

Sam Burrett
Consulting Lead, AI Advisory

What the data tells us

Theme 2:

AI is enterprise infrastructure, and governance maturity must catch up

98% of organisations have adopted AI platforms in the past 24 months.

95% have a formal AI governance framework.

In 2026,

64%

are 'very confident' in their controls to manage AI-related risks, compared with

36%

in 2025 results.

Twenty-four months ago, generative AI was only an emerging technology, principally in pilot. Our 2026 survey shows it is now the default enterprise infrastructure across the surveyed population. The statistics confirm broad adoption and formal governance coverage of AI, but they mask a maturity gap. As we discuss in detail in our AI Governance Guide, an AI policy is not AI governance.

Regulators have reached the same conclusion. On 22 June 2026, the Five Eyes cyber security agencies published a joint call to action, noting that '*AI is rapidly transforming cyber risk*' and calling for urgent action.[13] In its letter published on 30 April 2026,[14] APRA found that, while AI is being '*actively adopted by all the entities*' they engaged with, there are '*differing levels of maturity across functions such as governance, risk management and operational resilience*' and that '*assurance practices are not keeping pace with the scale, speed and complexity of AI*'.

The message from regulators is consistent: having a framework on paper does not mean adequate and resilient controls are in place.

In this rapidly evolving environment, our clients are increasingly commissioning AI Governance Maturity Assessments to identify where controls fall short of regulator expectations and to strengthen their position before – not after – a cyber incident or regulatory inquiry.

The survey data reinforces why governance must catch up. When respondents are asked to rank the risks of AI adoption itself, cyber security now leads.

41% rank cyber security as their number 1 AI risk, ahead of privacy and data protection **(23%)**, operational and model failure **(17%)**, and legal and regulatory risk **(14%)**.

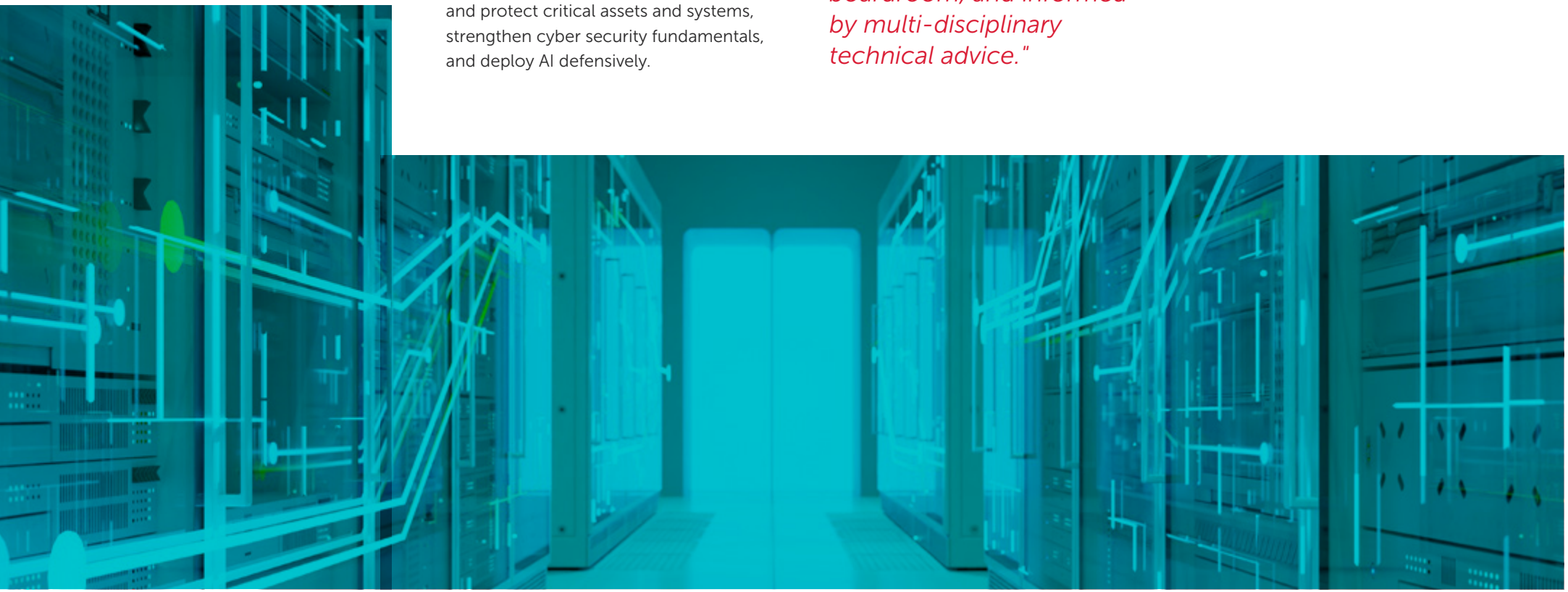
In last year's survey, privacy led the AI risk hierarchy. Cyber security has now displaced it at the top.

This convergence of cyber and AI risk is consistent with what we are observing in practice: an increasing proportion of cyber incidents and notifiable data breaches have a distinct AI dimension – whether as the vector, the target, or both.

The implication of this is that AI risks – cyber risks foremost among them – must be intentionally governed, not left to existing frameworks that were designed before AI became infrastructure. Boards and executives must allocate budget for AI investment and governance in concert, maintaining adequate oversight of the unique risks AI introduces. ASIC has called on entities to urgently reassess cyber plans, confirm cyber risk governance and decision-making frameworks, identify and protect critical assets and systems, strengthen cyber security fundamentals, and deploy AI defensively.

"New AI technologies and increased adoption materially expand the attack surface and create the risk of exposure at scale. AI governance and legal risk management is an enterprise-wide issue. It must be led from the boardroom, and informed by multi-disciplinary technical advice."

Chelsea Gordon
Legal Lead – AI Advisory



What this means for leaders in 2026

A written AI governance framework is the floor, not the ceiling. The substantive question is whether that framework is operationalised end-to-end and subject to ongoing monitoring and adaptation: a maintained AI register, risk-based use-case approvals, data and privacy controls embedded across the lifecycle, accountability assigned at executive level, third-party AI risk managed contractually, monitoring designed for memorisation and prompt-injection risks, and human oversight calibrated to the risk profile of each use case.

In an environment where regulators are technology-agnostic, the burden of demonstrating that existing duties have been properly applied to AI sits squarely with the organisation, with appropriate oversight and investment by the Board. This is the central message of the APRA and ASIC letters discussed in Themes 1 and 2, and is explored in detail in our AI Governance Guide: An AI policy is not AI governance.

Australia's regulatory position: technology-neutral, to targeted mandatory standards

The Australian Government's December 2025 National AI Plan confirmed it would rely on existing technology-neutral laws rather than introduce AI-specific mandatory guardrails. That position is now shifting. On 15 July 2026, the Prime Minister announced a set of mandatory Australian Standards for AI, established an Office of AI within the Department of the Prime Minister and Cabinet, and flagged legislation for introduction to Parliament in early 2027, subject to National Cabinet agreement in August 2026. The mandatory obligations foreshadowed to date are targeted rather than economy-wide, covering large data centres (energy, grid connection and water requirements) and copyright protection for creative works; the Government has said it does not intend to legislate 'for every possible eventuality or risk'. For Boards and executives, the regulatory task still remains one of mapping obligations across a patchwork of existing laws and standards, now with an emerging mandatory layer whose scope and timing are still to be settled.

Pending that detail, the Government's approach otherwise remains two-pronged: uplift and clarify existing technology-neutral laws (privacy, consumer, sectoral, workplace, anti-discrimination, financial services), and provide voluntary guidance through the Guidance for AI Adoption: 6 Essential Practices and sector-specific standards. Organisations that have not yet mapped how these overlapping obligations apply to their AI systems face increasing regulatory and litigation risk, and should now also track the design of the proposed Australian Standards for AI as it develops over the next 12 months.

Privacy at the centre

Privacy sits at the centre of AI governance and is an increasingly important dimension of AI cyber incident readiness. AI systems run on personal information and increasingly generate or infer more of it, and each of those touchpoints is regulated. The OAIC has been explicit that the *Privacy Act 1988* (Cth) (**Privacy Act**) governs the collection, use, disclosure and handling of personal information by AI across the entire lifecycle, from training and fine-tuning through to deployment and output, and that the information an AI system infers or generates about an identifiable individual is itself 'personal information' governed by the Australian Privacy Principles (**APPs**).

The OAIC's two guides of 21 October 2024 remain the baseline reference: one for organisations deploying commercially available AI products, and one for developers training or fine-tuning generative AI models.[16] Together they set a clear expectation of a governance-first, privacy-by-design approach across the AI lifecycle, including a privacy impact assessment before a new AI system is deployed, close attention to APP 3, where an AI system collects, generates or infers personal information, and to APP 10, where probabilistic outputs may be inaccurate. Both OAIC guides remain essential reference points for any organisation developing or deploying AI.

The privacy exposure is no longer only regulatory. Since 10 June 2025, individuals have been able to sue directly for serious invasions of privacy under the statutory tort in Schedule 2 of the Privacy Act. The tort covers both intrusion upon seclusion and misuse of information, is actionable without proof of damage, and reaches defendants well beyond APP entities. For organisations deploying AI-enabled surveillance, profiling or large-scale data collection, this is a materially new litigation risk, and one that can crystallise alongside regulatory action in the aftermath of a cyber incident.

The OAIC has also published specific guidance on facial recognition technology (**FRT**), treating live FRT, including where AI-enabled, as highly intrusive and expecting organisations to conduct a privacy impact assessment, adopt privacy by design, and apply the APPs rigorously.[17]

The determinations discussed below demonstrate how these regulatory expectations have been applied is in practice.

A further obligation is close at hand. From 10 December 2026, new APPs 1.7 to 1.9 will require APP entities to disclose, in their privacy policies, where they use personal information in a computer program to make, or substantially assist in making, decisions that could reasonably be expected to significantly affect an individual's rights or interests.
The OAIC has begun a compliance sweep of privacy policies and will publish guidance before commencement. In the meantime, organisations should already be mapping their automated decision-making and auditing what their policies will need to disclose by that date. (Explore more in our summary: Automated decision-making obligations for businesses).

The connection to cyber risk is direct. The more an organisation relies on AI, the more personal information it holds and creates, enlarging both its APP 11 security obligations and the volume of data that may be exposed when an incident occurs. Privacy governance and cyber resilience, in this context, are not separate workstreams; but address the same underlying exposure.

Recent enforcement and what it signals

Recent Australian regulatory and judicial activity demonstrates how existing technology-neutral obligations are being applied to AI-enabled systems in practice. For organisations managing cyber risk, these cases are directly relevant: each involves the kind of data collection, automated processing or surveillance technology that sits within the modern cyber-attack surface, and each illustrates how regulatory enforcement can follow where governance, transparency or proportionality falls short. The cases below illustrate why judicious legal and AI risk assessment, including privacy impact assessments, must precede approval of new AI use cases.

Bunnings[18]

Between November 2018 and November 2021, Bunnings deployed FRT across up to 62 stores to address repeated serious incidents of assault, harassment, robbery and theft by repeat offenders. The system captured real-time facial images of the offenders via CCTV, converted them into vector sets (mathematical representations of facial geometry), and compared these against a database of enrolled individuals.

The Privacy Commissioner determined in October 2024 that Bunnings' collection of biometric data breached APP 3.3. Bunnings applied to the Administrative Review Tribunal (**ART**) for review, and in February 2026 the ART overturned the APP 3.3 finding on the basis that Bunnings could rely on the permitted general situation exceptions, due to their unique security environment, the seriousness of the criminal activity, and the privacy-by-design features of its FRT system (including near-immediate deletion of unmatched data).

The ART nonetheless affirmed breaches of APPs 5.1, 1.2 and 1.3 for inadequate notification, governance and privacy policy disclosure. For organisations managing cyber and data risk, the decision confirms that even momentary capture of facial images constitutes 'collection' of sensitive information, that biometric systems must be accompanied by documented governance and transparency measures, and that the threshold for justifying deployment of privacy-intrusive technology remains high. The decision is not a green light for widespread FRT deployment.

Kmart[19]

Between June 2020 and July 2022, Kmart deployed FRT in 28 of its retail stores, capturing the face of every person who entered and presented at returns counters. These images were processed to generate biometric templates cross-referenced against a database of individuals suspected of engaging in fraudulent returns.
In September 2025, the Privacy Commissioner determined that Kmart's collection was unlawful. Unlike Bunnings, Kmart could not establish the permitted general situation exception: the collection was indiscriminate (capturing every customer, not only those who triggered concern), of limited utility, and disproportionate to the harm Kmart sought to prevent. Less-intrusive measures requiring proof of purchase, redesigning store layouts, enhancing staff training – were available and had not been adequately considered.

The Privacy Commissioner found that capturing the biometric data of tens of thousands of customers to address a relatively small number of fraudulent cases was disproportionate, and that the risks of mass biometric surveillance including misuse, discrimination and erosion of privacy, outweighed the benefits.

Kmart was directed to cease using FRT, publish a public statement and apology, and ensure deletion of retained biometric data. For organisations deploying AI-enabled surveillance or automated data collection systems, the determination reinforces that indiscriminate collection of sensitive information – even where a genuine business problem exists – will not satisfy the proportionality requirements of the Privacy Act, and that the availability of technology does not itself justify its use.

Clearview AI[20]

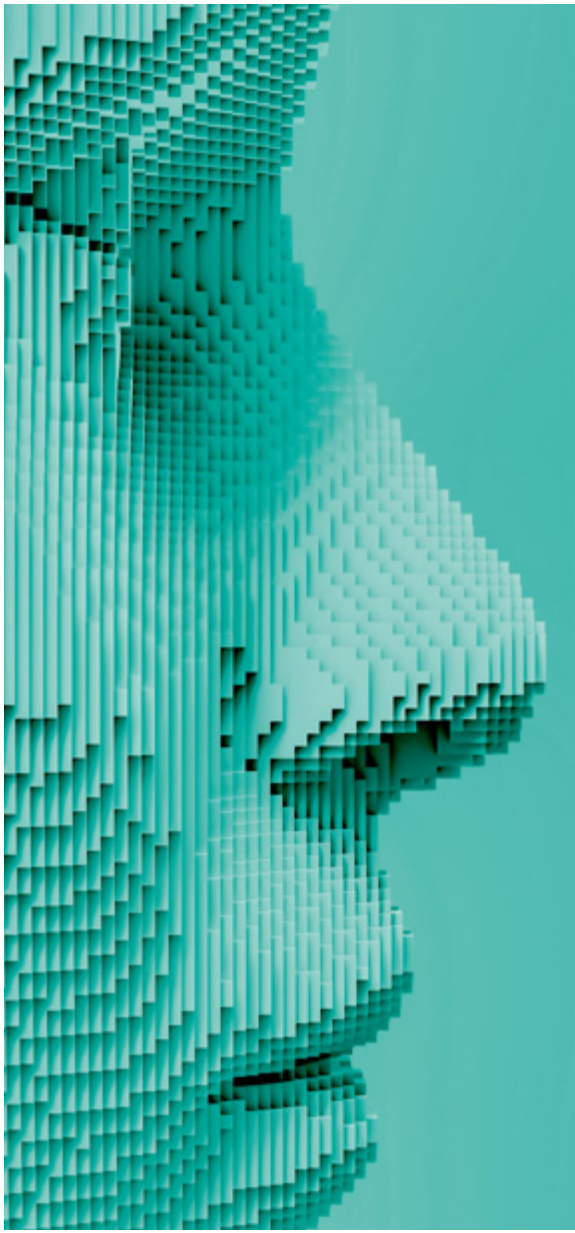
Following a joint investigation between the OAIC and the United Kingdom Information Commissioner's Office, the Administrative Appeals Tribunal (**AAT**) upheld the OAIC's determination that Clearview AI's scraping of facial images from publicly available web sources to populate a commercial facial recognition database breached the Privacy Act. The AAT confirmed that Clearview AI was '*carrying on a business in Australia*' by virtue of its repeated, automated collection of images from Australian servers even though it had no physical presence in Australia and derived no revenue from Australian operations.

The decision established that publicly accessible does not equal lawfully usable: the fact that images were made publicly available on social media did not authorise their collection for a different commercial purpose. Critically, the AAT also confirmed the expanded extraterritorial reach of the Privacy Act following the 2022 amendments; once an Australian link is established, a foreign corporation's global data handling practices fall within the Privacy Act's scope.

For organisations managing cyber and AI risk, the decision is directly relevant to generative AI training data sourced through web scraping, to the use of third-party AI tools whose training data provenance is uncertain, and to any automated system that collects personal information at scale from publicly available sources. It also underscores the regulatory risk for organisations using offshore AI vendors whose data practices may independently trigger Australian privacy obligations.

The takeaways from these cases connect directly to theme three's premise: AI is enterprise infrastructure, and governance must catch up. Each of these cases involved automated or AI-enabled data collection at scale, and each resulted in regulatory enforcement because governance, transparency or proportionality was found wanting. For organisations managing cyber risk, the lesson is that the same systems that create operational value also create regulatory and litigation exposure. This exposure may crystallise rapidly in the event of a cyber incident, when regulators, affected individuals and class action plaintiffs will scrutinise not only how data was secured, but whether it should have been collected at all, whether individuals were adequately informed, and whether the organisation's governance was defensible.

For organisations, this means mapping their AI obligations, performing appropriate risk assessments of new technology and AI systems, and ensuring mature, operationalised governance is in place – both to manage AI safely in the ordinary course and to withstand scrutiny in the event of a cyber incident.



What the data tells us

Theme 3:

Preparing for the wrong incident?

For more than a decade, cyber preparedness has been measured in our surveys by reference to two indicators: whether organisations have an incident response plan, and whether they test it. On both metrics, progress has been substantial.

As of 2026, that maturity appears to have plateaued with 91% of surveyed organisations having an incident response plan and 93% testing it at least annually. What was once a differentiator is now baseline.

The question now is not whether organisations are prepared, but what they are prepared for.

Plan adoption has risen from **42%** in 2016 to over **91%** in 2026, while regular testing has increased from **34%** to over **70%** across this period.

Scenario libraries have not caught up to the threat

Most contemporary incident response playbooks are built on a recognisable family of scenarios: ransomware, business email compromise, insider misuse, supply-chain attacks and denial of service. These threats remain real, but two things have changed.

First, as discussed in Theme 1, the threat surface has widened to take in a family of AI-era scenarios, deepfake-enabled fraud, prompt-injection compromise, agentic offensive activity, that few playbooks yet rehearse, even though AI-enabled threats are now organisations' second-highest named concern.

Second, some familiar risks, supply-chain compromise chief among them, have grown in reach in a way the standard tabletop exercise may not fully address.

For large, higher-risk and more digitally dependent organisations, the rehearsal cycle should now include the following threat vectors:

Deepfake-enabled fraud, including voice-cloned vishing of finance teams

- The reference case remains the 2024 Arup matter, in which an employee transferred approximately US$25 million after joining a video conference whose other participants were all deepfaked colleagues.[21]

Prompt-injection compromise of enterprise AI tools

- The EchoLeak vulnerability, disclosed in June 2025, demonstrated that data could be exfiltrated from Microsoft 365 Copilot through a single crafted email, without the recipient clicking a link or opening an attachment.[22]

Supply-chain compromise, where the breach sits with a vendor and the organisation is downstream of it

- Two aspects tend to go unrehearsed: where the vector is a service provider's AI system rather than the organisation's own, and the response itself, which may mean reaching affected individuals with whom the organisation has no relationship, and coordinating across other entities in the chain that each carry their own reporting obligations.

Synthetic-media reputational events

- This includes circumstances where fabricated audio, video or documents are used to destabilise market, customer, regulator or workforce responses in the early hours of an incident.

None of this is hypothetical any longer. A playbook drawn up before mid-2024 is unlikely to deal with any of these scenarios well.

Response is now scrutinised through a sharper legal lens

The scenario library is only part of the problem. Even where a plan anticipates the right incident, the legal and regulatory environment around a cyber incident has changed substantially over the past decade. This Report has tracked that shift as it has unfolded:

- 2018 – The Notifiable Data Breaches scheme commenced under the Privacy Act, introducing mandatory notification to the OAIC and affected individuals for breaches likely to cause serious harm.
- 2019 – APRA's CPS 234 Information Security takes effect, requiring APRA-regulated entities to maintain baseline security controls and third party assurance.
- 2021–2022 – The Security Legislation Amendment (Critical Infrastructure) Acts extended the SOCI Act from four to 11 sectors, introducing mandatory incident reporting and, for some entities, critical infrastructure risk management programs.
- 2022 – In ASIC v RI Advice Group, the Federal Court found a breach of licensee obligations under section 912A of the Corporations Act for inadequate cyber security controls, ASIC's first cyber enforcement action and an early signal that cyber risk is a director's duty.
- December 2022 – Following the Optus and Medibank breaches, maximum civil penalties for serious privacy interferences rose from $2.22 million to the greater of $50 million, three times the benefit obtained, or 30% of adjusted turnover.
- November 2024 – The Cyber Security Act 2024, Australia's first standalone cyber security legislation, introduced mandatory ransomware payment reporting and established a Cyber Incident Review Board to conduct post-incident reviews of significant events.

- December 2024 – The Privacy and Other Legislation Amendment Act 2024 lowered the top civil penalty threshold (a breach need only be serious, not serious and repeated), introduced a new mid-tier penalty, and laid the groundwork for the statutory tort and automated decision-making transparency obligations discussed below.
- March 2025 – ASIC brought its second cyber enforcement action, against FIIG Securities, reinforcing that cyber risk management is a legal obligation rather than a discretionary good practice.
- April 2025 – Telecommunications security obligations were consolidated into the SOCI Act, and the definition of critical infrastructure was extended to data storage systems holding business-critical data.
- July 2025 – APRA's CPS 230 Operational Risk Management commenced, raising the bar on third-party and material service provider risk management.
- Ongoing — The OAIC has shifted from guidance to civil penalty litigation, with proceedings against Meta, Medibank and Australian Clinical Labs.

Two developments since our 2025 edition are directly relevant here, though in different ways.

The new **statutory tort for serious invasions of privacy** changes what organisations need to have already done before an incident occurs. **Legal professional privilege** changes how the response itself needs to be structured once one does.

The statutory tort for serious invasions of privacy

The tort, introduced by the Privacy and Other Legislation Amendment Act 2024 (Cth) and in force since 10 June 2025, applies to serious invasions of privacy committed intentionally or recklessly, by intrusion upon seclusion or misuse of private information, where the public interest in privacy outweighs any countervailing public interest. Damages for non-economic loss, together with any exemplary or punitive award, are capped at the greater of AUD$478,550 and the maximum available for non-economic loss in defamation. Proven economic loss sits outside this cap.

For cyber incidents, the significance of the tort is less about how a response plan is drafted and more about what an organisation can show it had already done before the incident happened. The tort covers reckless conduct, not just intentional conduct, meaning inadequate data governance, weak access controls, or a known but unaddressed vulnerability can each found a claim. It confers a direct cause of action, reaching beyond APP entities, without requiring the individual to proceed through the Privacy Commissioner first, and the one year limitation period from the plaintiff's awareness of the invasion places a premium on how quickly an organisation can identify and respond to exposure.

The practical implication of the tort is that post-incident litigation risk now turns substantially on pre-incident conduct – what data an organisation held, why it still held it, whether that holding was proportionate, and whether known gaps had been addressed. That is a data governance and risk-assessment exercise, not solely a response-plan exercise, and it needs to be substantially complete before an incident occurs.

In *McClure v Medibank Private Limited* [2025] FCA 167 (**McClure**), the Federal Court of Australia held that three Deloitte reports commissioned after Medibank's 2022 breach were not protected by legal professional privilege. In April 2026, the Full Federal Court (Wigney, Lee and Hespe JJ) refused Medibank's leave to appeal.[26]

The Full Federal Court's findings matter for anyone planning an incident response. For privilege to attach to a post-incident report, the legal purpose must be the **dominant** one, assessed objectively against the organisation's institutional purpose, rather than the stated belief of any individual executive. That distinction is what makes a cyber response so exposed.

A serious incident is handled at once by the Board, a crisis committee, the executive, external affairs, risk and governance, technical advisers and the legal team, and the more genuinely cross-functional the response, the harder it becomes to show that obtaining legal advice prevailed over every other purpose.

In Medibank's case, the engagement of Deloitte through its lawyers helped but did not decide the question; what counted against the claim was the surrounding conduct, including a public ASX statement that the firm had been engaged to '*learn from this event*', and active engagement with the regulator over the review's terms.

The practical consequence of *McClure* is that the architecture of a post-incident engagement cannot be improvised once an incident is underway. It has to be settled in advance and built into the incident response plan itself. In practice, that means settling – *before* an incident occurs – who commissions sensitive reviews and through which lawyers, recording the legal purpose of each engagement, keeping it genuinely dominant rather than nominal, and holding a consistent line in public and with regulators about why a review is being run.

None of this can be reconstructed convincingly after the fact. A plan that rehearses the technical response, but leaves the legal architecture to be assembled in the first hours of a crisis, is built without regard to the lessons of *McClure*.

The three incidents discussed below demonstrate why this matters in practice.

A stale playbook can fail for reasons that have nothing to do with AI: the breach sits in the supply chain, the entry point is a known edge-device vulnerability, or the organisation cannot quickly explain what data it held, why it held it, how it was secured, and who was affected. AI does not remove these problems. But it just makes them faster to occur and harder to contain.

INCIDENT 1 ›

youX: the supply-chain scenario, no longer hypothetical

In February 2026, the Sydney asset-finance platform youX confirmed that an unsecured cloud database had exposed the personal and financial records of 444,538 Australian borrowers, including driver's licences, residential addresses and loan-application data, after a threat actor released a sample of the stolen data and demanded a ransom.[27] youX notified the OAIC, but the defining feature of the incident was its reach.

A single compromise at one vendor created notification and downstream-risk obligations across some 800 broker firms and more than 90 lenders, and drew at least one listed company into ASX disclosure over its exposure.

This scenario exposes exactly the parts of the response that tend to go unrehearsed: reaching affected individuals the organisation has no relationship with, and coordinating across other entities in the chain that each carry their own reporting obligations. A tabletop exercise built around a direct compromise of the organisation's own systems will not prepare for this.

INCIDENT 2 ›

Ivanti VPN vulnerabilities: existing risks have not gone away (and are now heightened)

Vulnerabilities in widely deployed Ivanti remote access products were actively exploited through 2024 and into 2025, including for remote code execution, persistence and anti-forensic techniques. ASD's Annual Cyber Threat Report 2024-25 recorded more than 120 incidents associated with edge-device attacks during the year, of which 96% succeeded.[28]

This incident highlights a different point: widening the rehearsal cycle to take in AI-era threats is necessary, but these are additions, not replacements. Familiar vulnerabilities remain just as dangerous, and AI is now accelerating attacks on them too.

AI now does to older vulnerabilities what it does to new ones: finding vulnerable internet-facing assets, reading code, identifying exploitable flaws and generating working exploits.

The result is a shrinking patch window. A known, unpatched vulnerability in an edge device was already one of the most reliable ways into an organisation; AI makes that pathway faster, cheaper and harder to close.

A current playbook therefore needs to test not only whether the organisation patches, but how quickly it can identify exposure, make risk-based containment decisions, communicate with affected stakeholders, and justify those decisions afterwards.

INCIDENT 3 ›

MediSecure: the exposure is set before the incident

The compromise of the e-prescription provider, which occurred in May 2024, exposed personal and health information relating to approximately 12.9 million Australians, the largest breach ever notified to the OAIC under the Notifiable Data Breach scheme.[29] The matter drew sustained government involvement, including coordination through the National Cyber Security Coordinator, and the company entered into administration the following month after the Federal Government declined to fund its continued operation.

MediSecure is a clear illustration that legal exposure may be set well before the incident occurs. Where sensitive personal information is involved, attention moves straight to the data itself: what the organisation held, why it still held it, how long it had kept it, and how it was secured. The statutory tort raises the stakes further, since affected individuals now have a direct cause of action that doesn't depend on the regulator acting first. A breach exercise that starts with containment but has not tested supply chain, data-mapping, retention, notification and affected-individual analysis starts too late.

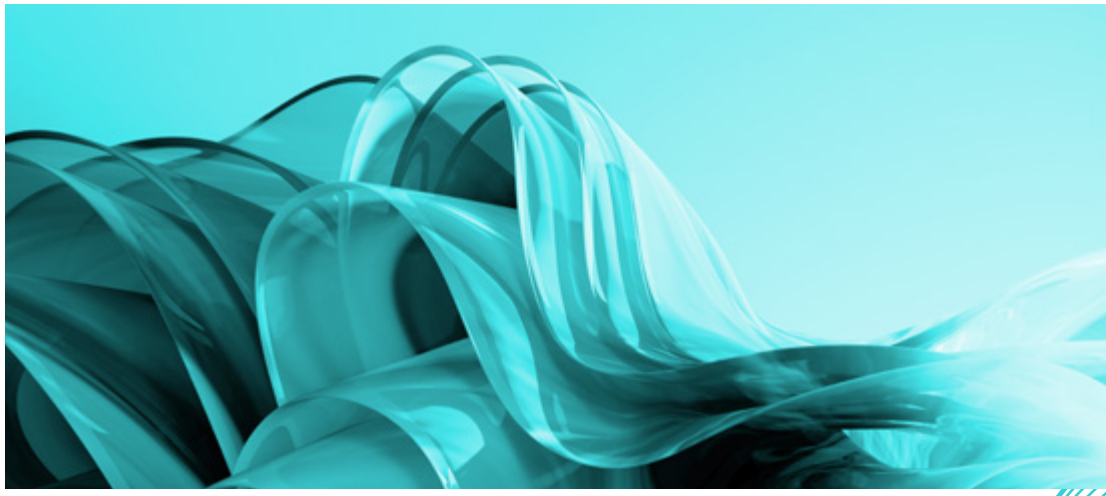
"Almost every organisation in our survey now has a plan and tests it. But fewer have asked the harder question: have we rehearsed the incident we are actually going to get, or just the ones we already know how to handle?"

Paul Kallenbach
Partner, Technology, Digital & Data; National Legal Cyber Leader

Three questions follow from this theme, and Boards should ask each of them directly:

- Does the playbook contemplate the incidents the organisation is genuinely likely to face: both the AI-era scenarios it ranks among its leading concerns, and the 'traditional' risks, such as supply-chain compromise, that may now have far-reaching impacts that remain unrehearsed?
- Has the legal shape of the response been settled in advance and written into the plan, rather than improvised in the first hours of a crisis? This extends well beyond privilege. To overlap regulatory reporting clocks; under the statutory tort and data breach class actions that may follow; continuous-disclosure obligations for listed entities once an incident is material; the 'reasonable steps' expected under APP; contractual notification duties owed up and down the supply chain; and directors' own duties of oversight.
- Does the exercise reach beyond the technical team to the legal, communications and regulatory affairs? Third parties and affected individuals who sit outside the organisation's own systems built into the plan, rather than discovered in the midst of the crisis?

A playbook that rehearses yesterday's breach offers little guidance when the next one is AI-enabled, supply-chain driven, and moving at machine-speed.



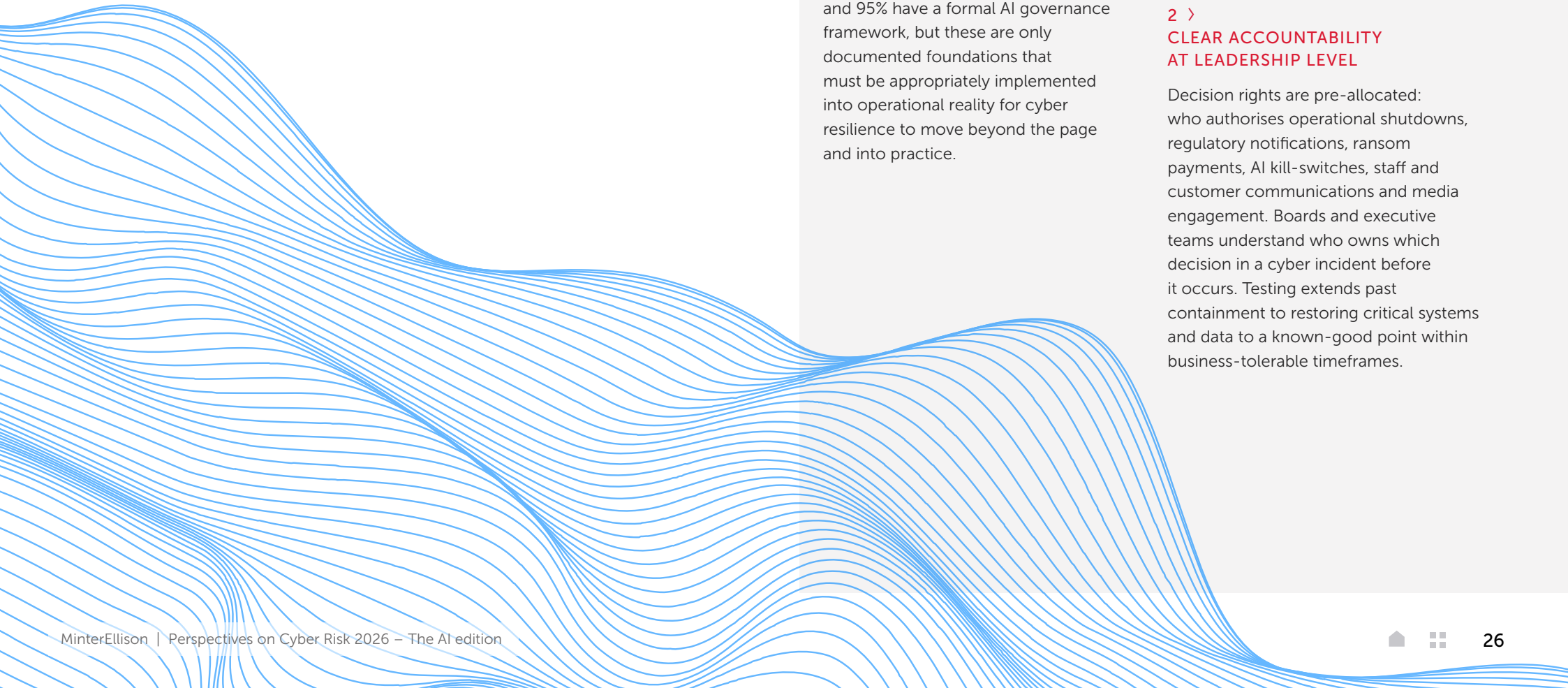
Where to from here: priorities for Boards

Where to from here: priorities for Boards

"Preparedness is not a static state; it is an ongoing process of testing, learning and adapting, and must be led from the boardroom."

Paul Kallenbach
Partner, Technology, Digital & Data;
National Legal Cyber Leader

Cyber resilience in 2026 is defined by how effectively an organisation can anticipate, respond, recover and adapt when disruption occurs, and by whether the decisions it made during the response can be shown, both before and after the fact, to have been timely, informed and defensible.



Leading organisations share eight characteristics.

They are simple to describe but demanding to demonstrate because they must be operationalised to be effective.

91% of surveyed organisations have an incident response plan, and 95% have a formal AI governance framework, but these are only documented foundations that must be appropriately implemented into operational reality for cyber resilience to move beyond the page and into practice.

1 ›
CLEAR VISIBILITY AT BOARD LEVEL

The Board has direct line of sight over the organisation's cyber readiness and AI governance strategy, aligned with organisational risk appetite, with these items regularly appearing on the Board's agenda for review and discussion.

2 ›
CLEAR ACCOUNTABILITY AT LEADERSHIP LEVEL

Decision rights are pre-allocated: who authorises operational shutdowns, regulatory notifications, ransom payments, AI kill-switches, staff and customer communications and media engagement. Boards and executive teams understand who owns which decision in a cyber incident before it occurs. Testing extends past containment to restoring critical systems and data to a known-good point within business-tolerable timeframes.

3 ›
TESTED RECOVERY AND RESPONSE CAPABILITY ACROSS THE RIGHT SCENARIOS

Plans are exercised against AI-era scenarios, including deepfake-enabled fraud, prompt-injection of internal AI systems, supplier-platform compromise and agentic offensive activity, not just ransomware, business email compromise and denial of service.

4 ›
JOINED-UP DECISION-MAKING

Legal, technology, risk, communications and operational teams are brought together early, with clear escalation paths and a shared operating picture to ensure that no single function drives the response in isolation.

5 ›
VISIBILITY ACROSS CRITICAL EXPOSURES

The organisation maintains a current, classified inventory of critical systems (including AI systems deployed across the supply chain), data flows, sensitive data holdings, privileged-access pathways and material service providers, at the level CPS 230 expects of regulated entities, and increasingly at the level the OAIC, ASIC, ACCC and class action plaintiffs expect of all organisations.

6 ›
CYBER FUNDAMENTALS ARE UPLIFTED FOR MACHINE SPEED

The basics, including patching, validating core controls, minimising the attack surface, and tightly managing privileged access, are done well and continuously. These foundations are resourced and maintained at the pace an AI-augmented threat environment now demands: a priority ASIC and APRA have both stressed this year.

7 ›
AI IS PUT TO WORK IN DEFENCE

AI is used to strengthen and defend, identifying vulnerabilities at scale; triaging alerts; and supporting patching, enabling the organisation to match the speed of its adversaries.

8 ›
CONTINUOUS IMPROVEMENT

Lessons from exercises, real incidents, regulator activity, horizon scans and technology change are translated into stronger practical controls, clearer governance and faster response, not just into a refreshed policy document. An AI policy is not AI governance.

Recommended Board and executive priorities for the year ahead:

1 ›
STRESS-TEST YOUR RESPONSE PLAN AGAINST AI-ERA SCENARIOS

If the last tabletop exercise did not include at least one deepfake-enabled fraud scenario (the Arup incident, referred to in Themes 1 and 3, remains the reference case), one prompt-injection scenario (such as the EchoLeak vulnerability disclosed in June 2025), and one supplier-platform compromise scenario, schedule this now.

Bring in legal, communications and a Board-level participant, as the speed of executive decision-making is where most failures occur. Assess how your Cyber Incident Response Plan and AI Governance Framework assist with this scenario planning and where changes are needed.

2 ›
CLARIFY DECISION RIGHTS IN ADVANCE

Ask:

- Who performs the notifiable data breach assessment?
- Who authorises regulator notifications within the SOCI Act, ransomware, CPS 234, notifiable data breach and other regulatory reporting windows?
- Who decides on a public statement before forensic facts are known, and what can be said publicly without waiving privilege?

Define these in writing and ensure owners are aware of their responsibilities.

3 ›
REASSESS THIRD-PARTY RISK ON A CPS 230 FOOTING

The youX incident discussed in Theme 3 illustrates why a single compromise at one vendor created notification and downstream-risk obligations across some 800 broker firms and more than 90 lenders. Whether or not your entity is APRA-regulated, CPS 230 provides a credible framework for managing third-party risk (noting that most AI risk sits within the supply chain): classify material service providers, document the operational risk they introduce, embed contractual controls (notification, audit, sub-processor restrictions, AI-use disclosure), and test continuity for severe-but-plausible failure scenarios. For critical AI dependencies, add sovereign access risk to the analysis, as we have now seen that a single offshore frontier model can be withdrawn by a foreign regulatory decision. Map single-provider reliance and hold a tested fallback. Explore more in our Third-party contracting for AI article.

4 ›
OPERATIONALISE AI GOVERNANCE AND EMBED ASSURANCE

95% of respondents have a formal AI governance framework – but as APRA observed in its April 2026 letter, '*assurance practices are not keeping pace with the scale, speed and complexity of AI*'. The 2026 test is whether the framework is operating in practice in the form of an AI Management System, with key relevant components such as a maintained AI register; risk and impact assessments; documented use-case approvals; data, privacy and IP controls embedded; monitoring designed for AI system attacks such as prompt-injection and memorisation; clear human-oversight requirements at the AI system and use case levels; and AI procurement and vendor risks integrated and managed.

Given both ASIC's and APRA's concerns about assurance, Boards should seek appropriate internal and external assurance that AI governance practices have been validated against a recognised standard. ISO/IEC 42001:2023 is the recognised international standard for AI management systems, and can be overlaid with the Government's Guidance for AI Adoption and applicable sectoral standards.[30]

5 ›
DESIGN LEGAL PRIVILEGE INTO YOUR CYBER RESPONSE ARCHITECTURE FROM DAY ONE

As the discussion of *McClure* in Theme 3 makes clear, the architecture of a post-incident engagement cannot be improvised once an incident is underway. Engagement structures, scoping of expert reports, and what is said publicly about what is being investigated must be settled in advance with a clear dominant legal purpose. It is increasingly common for organisations to brief separate experts for the data breach response and for any legal analysis, so that the legal expert's report is protected by legal professional privilege.

Record the legal purpose of each engagement, keep it genuinely dominant rather than nominal, and hold a consistent line, in public and with regulators, about why a review is being run.

6 ›
ALIGN CYBER READINESS WITH THE NEW REPORTING AND DISCLOSURE OBLIGATIONS

Key new obligations include:

- Mandatory ransomware payment reporting under the *Cyber Security Act* 2024 (Cth) (live since 30 May 2025; enforcement phase from 1 January 2026).
- The statutory tort for serious invasions of privacy, which confers on affected individuals a direct cause of action without requiring them to proceed through the OAIC (live since 10 June 2025).
- APRA CPS 230 (live since 1 July 2025).
- Telecommunications security obligations consolidated into the SOCI Act (new Part 2D), and data storage systems holding business-critical data brought within critical infrastructure risk-management obligations (live since 4 April 2025).
- APP 1 automated decision-making transparency obligations (from 10 December 2026).

7 ›
REVIEW CYBER INSURANCE AGAINST AI-ERA EXPOSURES

94% of surveyed organisations hold cyber insurance, but few policies were drafted with autonomous AI-enabled attacks in mind. Ask your broker or insurer directly: does the policy respond where an autonomous AI agent, rather than a human operator, executes the intrusion, and could an insurer dispute causation or attribution on that basis? Do war, state-actor or terrorism exclusions risk being invoked more readily against frontier-model-enabled attacks than they were against human-directed ones? Does cover extend to incidents originating in a third-party AI platform your organisation uses but does not control?

Review these questions now, before a claim tests them. Insurers are still developing their own view of AI-related risk, and policy wording in this area is moving quickly; a policy reviewed even 12 months ago may no longer reflect current exclusions or sub-limits.

8 ›
BUILD BOARD LITERACY IN CYBER AND AI

Clear visibility at Board level, discussed above, depends on the Board having the literacy to act on what it sees. The Diligent Institute's APAC Governance Outlook report, published in November 2025 in partnership with the Governance Institute of Australia and the Singapore Institute of Directors, found that only 13% of Australian boards have recruited a director with AI expertise, and just 21% mandate director training on AI, even as 43% of Australian governance leaders rank AI adoption as their top strategic priority, and 98% of our surveyed organisations told us they adopted AI in the last 12 months. Cybersecurity fares little better as a Board capability question: 53% of Australian boards name it their top strategic priority, well above the APAC regional average, but priority is not the same as literacy.

Appoint at least one director with genuine cyber or AI expertise, or engage an independent expert to brief the Board regularly, not only after an incident. Ensure directors can ask informed questions about AI system risk, model behaviour and assurance, not only about traditional network security, since AI systems are now a significant source of the organisation's cyber exposure.

Each adds a discrete dimension to incident response or compliance posture that needs to be exercised, not just documented. In particular, the combination of the statutory tort and the privacy enforcement cases discussed in Theme 2 means that data minimisation and retention review are no longer merely good practice, but will directly reduce the volume of data exposed should an incident occur.

Conclusion

In 2026, the measure of cyber resilience is not whether incidents can be prevented, but how effectively organisations respond when they occur. Across this year's findings, the pattern is clear: cyber incidents are now routine, AI has become embedded enterprise infrastructure, and the threat landscape is moving faster than the governance models designed to manage it.

While 91% of organisations now have an incident response plan and 93% test it, the focus has shifted from preparedness in form to preparedness in substance. The key question is this: has the organisation anticipated how incidents now manifest, and whether its response can stand up to legal, regulatory and stakeholder scrutiny?

The organisations best positioned in the year ahead will be those that recognise the shift created by an emerging AI-enabled world, and act on it. That means stress testing response plans against AI-era scenarios, operationalising AI and cyber governance frameworks so they function in practice, and embedding accountability, legal architecture and decision-making discipline before an incident occurs. In an environment where AI adoption has reached near universal levels and the attack surface continues to expand, resilience is no longer an operational capability, but a leadership one.

About MinterEllison and authors

MinterEllison is one of Australia's largest law firms, recognised for its work in technology, cyber, AI, privacy and data

Our cyber and AI Advisory specialists advise Boards, executive teams, general counsel, CISOs and government agencies on cyber risk strategy, incident response, regulatory engagement, privacy and data protection, AI governance and class action defence. We have led the legal response to many of the most significant cyber incidents to affect Australian organisations over the past decade.

For further information or to discuss the issues raised in this report, please contact our team.

Paul Kallenbach
Partner, National Legal Cyber Leader
E Paul.Kallenbach@minterellison.com

Chelsea Gordon
Legal Lead, AI Advisory
E Chelsea.Gordon@minterellison.com

Sam Burrett
Consulting Lead, AI Advisory
E Sam.Burrett@minterellison.com

Acknowledgements

Thank you to all those who contributed to the 11th edition of ***Perspectives on Cyber Risk Report 2026: The AI edition***.

- Caru Froneman
- Wanda Kuai
- Mika Glance

About the research

This report draws on findings from MinterEllison's 11th annual Perspectives on Cyber Risk survey. The 2026 survey was conducted in 2026 across 150 senior decision-makers from Australian organisations with 200 or more full-time equivalent employees, spanning banking, insurance, healthcare, professional services, technology, manufacturing, retail, energy, transport and government sectors.

Respondents included Board directors, C-suite executives, general counsel, CISOs, heads of risk and senior privacy and compliance leaders.

Where year-on-year comparisons are drawn against the 2025 (10th anniversary) edition or earlier editions, methodological differences in cohort composition, question wording and response options have been considered. Findings are reported as percentages of respondents who selected a given option; sample-size effects mean that small year-on-year movements should not be over-interpreted.

The legal, regulatory and incident commentary in this report is current as at August 2026 and is drawn from public sources, regulator publications, court judgments, government announcements and reputable industry reporting. Nothing in this report should be taken as legal advice in relation to any specific matter. Where readers wish to discuss any of the issues raised, the authors are available through the contact details section.

Endnotes

[1] Australian Cyber Security Centre, Annual Cyber Threat Report 2024–25 (Report, 14 October 2025) 35 : The $202,700 figure refers to the average cost per cybercrime incident for large Australian businesses in FY2024–25, a 219% YoY increase. Medium businesses experienced a 55% increase to $97,200 per incident.

[2] *Privacy and Other Legislation Amendment Act 2024* (Cth); the statutory tort for serious invasions of privacy is contained in Schedule 2 of the *Privacy Act 1988* (Cth) and commenced 10 June 2025. See Patrick Considine et al, 'Statutory Tort for Serious Invasions of Privacy Comes into Force', *MinterEllison* (Web Page, 11 June 2025)

[3] 'Final Targeted Amendments to CPS 230 Operational Risk Management', *Australian Prudential Regulation Authority* (Web Page, 30 April 2026): The APRA Prudential Standard commenced on 1 July 2025. APRA's targeted amendments to CPS 230 take effect from 1 July 2026.

[4] Department of Industry, Science and Resources, National AI Plan (Report, 2 December 2025); See Sam Burrett, Chelsea Gordon and Jason McQuillen, 'Australia Introduces a National AI Plan: Four Things Leaders Need to Know', *MinterEllison* (Web Page, 3 December 2025); 'Establishment of Australian AI Safety Institute', *Ministers for Department of Industry, Science and Resources* (Media Release, 24 November 2025): The Australian AI Safety Institute was jointly announced on 24 November 2025.

[5] Office of the Australian Information Commissioner, Automated Decision-Making Transparency Obligation (APP 1): Issues Paper (Report, 18 May 2026): Submissions to this issues Paper close 15 June 2026, with final guidance expected to be published by September 2026. The automated decision-making obligation was introduced by Schedule 1 of the *Privacy and Other Legislation Amendment Act 2024* (Cth) as new subclauses 1.7 to 1.9 of APP 1 of the *Privacy Act 1988* (Cth), commencing 10 December 2026. The Western Australian *Privacy and Responsible Information Sharing Act 2024* (WA) introduces a more onerous automated decision-making principle under IPP 10 for Western Australian-regulated entities from 1 July 2026.

[6] 'APRA Letter to Industry on Artificial Intelligence (AI)', *Australian Prudential Regulation Authority* (Web Page, 30 April 2026).

[7] 'Facial Recognition Technology: A Guide to Assessing the Privacy Risks', Office of the Australian Information Commissioner (Web Page, 19 November 2024).

[8] APRA Letter to Industry on Artificial Intelligence (AI) (n 6).

[9] 'ASIC Calls for Urgent Cyber Uplift as AI Accelerates Cyber Threats', *Australian Securities and Investments Commission* (Media Release, 8 May 2026).

[10] National Cyber Security Centre (UK), It's Time to Act: NCSC Annual Review 2025 (Report, 14 October 2025).

[11] Hong Kong police reporting on the Arup deepfake fraud, February 2024 (approximately HK$200 million / US$25.6 million transferred across multiple transactions following a deepfake video conference): Kathleen Magramo, 'British Engineering Giant Arup Revealed as $25 Million Deepfake Scam Victim', *CNN* (Web Page, 17 May 2024); Lo Hoi-ying, 'UK Multinational Arup Confirmed as Victim of HK$200 Million Deepfake Scam that Used Digital Version of CFO to Dupe Hong Kong Employee', *South China Morning Post* (Web Page, 17 May 2024).

[12] 'How We Hacked McKinsey's AI Platform', *CodeWall* (Web Page, 9 March 2026); 'AI vs AI: Agent Hacked McKinsey Chatbot and Gained Full Read-Write Access in Just Two Hours', *The Register* (Web Page, 9 March 2026), Rashmi Ramesh, 'Autonomous Agent Hacked McKinsey's AI in 2 Hours', *BankInfoSecurity* (Web Page, 13 March 2026): McKinsey patched the disclosed endpoints within 24 hours of responsible disclosure.

[13] 'Five Eyes Cyber Security Agencies Statement', *Australian Cyber Security Centre* (Web Page, 22 June 2026).

[14] APRA Letter to Industry on Artificial Intelligence (AI) (n 6); ASIC Calls for Urgent Cyber Uplift as AI Accelerates Cyber Threats (n 9).

[15] National AI Plan (n 4); 'Guidance for AI Adoption: Foundations', *National AI Centre* (Web Page, 21 October 2025).

[16] 'Guidance on Privacy and the Use of Commercially Available AI Products', *Office of the Australian Information Commissioner* (Web Page, October 2024); 'Guidance on Privacy and Developing and Training Generative AI Models', *Office of the Australian Information Commissioner* (Web Page, October 2024).

[17] Facial Recognition Technology: A Guide to Assessing the Privacy Risks (n 7).

[18] Commissioner Initiated Investigation into Bunnings Group Ltd (Privacy) [2024] AICmr 230 (29 October 2024); *Bunnings Group Limited and* Privacy Commissioner (Guidance and Appeals Panel) [2026] ARTA 130 (4 February 2026).

[19] Commissioner Initiated Investigation into Kmart Australia Limited (Privacy) [2025] AICmr 155 (26 August 2025); 'Kmart's Use of Facial Recognition to Tackle Refund Fraud Unlawful, Privacy Commissioner Finds', Office of the Australian Information Commissioner (Web Page, 18 September 2025).

[20] Clearview AI Inc and Australian Information Commissioner [2023] AATA 1069 (8 May 2023).

[21] 'British Engineering Giant Arup Revealed as $25 Million Deepfake Scam Victim' (n 11); 'UK Multinational Arup Confirmed as Victim of HK$200 Million Deepfake Scam that Used Digital Version of CFO to Dupe Hong Kong Employee' (n 11).

[22] Ravie Lakshmanan, 'Zero-Click AI Vulnerability Exposes Microsoft 365 Copilot Data Without User Interaction', The Hacker News (Web Page, 12 June 2025). 'CVE-2025-32711 Detail', *National Institute of Standards and Technology* (Web Page, 17 June 2026). The vulnerability was a zero-click indirect prompt injection enabling data exfiltration from Microsoft 365 Copilot; Microsoft confirmed it was remediated.

[23] *Cyber Security Act 2024* (Cth) pt 3. Royal Assent was obtained on 29 November 2024 and the mandatory ransomware payment reporting obligations commenced 30 May 2025, with the enforcement phase commencing from 1 January 2026. See Maria Rychkova, Natalie Adler, Paul Kallenbach, 'Pay and Tell: Mandatory Ransomware Payment Reporting Obligations in Force', *MinterEllison*.

[24] *Privacy Act 1988* (Cth) sch 2 s 11(5).

[25] Kurraba Group Pty Ltd & Anor v Williams [2025] NSWDC 396.

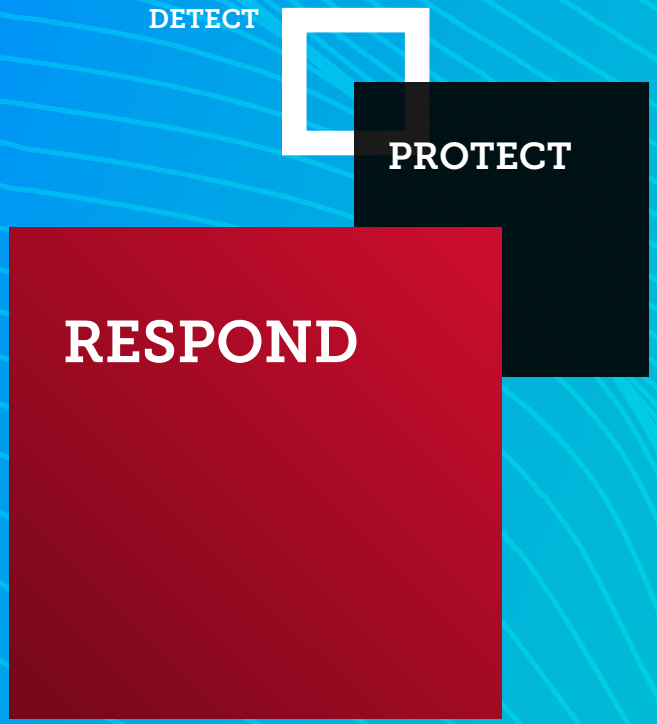
[26] McClure v Medibank Private Limited [2025] FCA 167. The appeal was refused in Medibank Private Limited v McClure [2026] FCAFC 38.

[27] 'youX Data Breach Exposes 444,000 Australians' Records', *Information & Data Manager* (Web Page, 20 February 2026); '19th February 2026 Cyber Update: youX Breach Exposes 444,000 Australians', *Cyber News Centre* (Web Page, 19 February 2026): The figure of 444,538 affected individuals is that confirmed by youX. Larger figures circulating in some reports, including loan-application counts and total data volumes, derive from the threat actor's claims and are not independently confirmed.

[28] See, eg, 'Critical Vulnerability in Pulse/Ivanti Connect Secure, Policy Secure and Neurons For ZTA Gateways (CVE-2025-22457)', *Australian Cyber Security Centre* (Web Page, 4 April 2025); Annual Cyber Threat Report 2024–25 (n 1) 35: The 84,700 cybercrime reports, 1,200+ incident responses (11% YoY increase), and 1,700+ proactive notifications (83% YoY increase) are headline figures from the ACSC's October 2025 release.

[29] 'Medisecure Cyber Security Incident', *Department of Home Affairs* (Web Page, 19 July 2024); 'Statement on MediSecure Data Breach', Office of the Australian Information Commissioner (Web Page, 13 September 2024): The figure of approximately 12.9 million being the largest number of affected individuals notified under the Notifiable Data Breaches scheme.

[30] Guidance for AI Adoption: Foundations (n 15).



minterellison.com

MinterEllison.