AI AND CYBER RESILIENCE

The Resilience Gap

The AI threat already inside your organsiation

Helen van Orton

Directorly Ltd

WHERE WE ACTUALLY ARE

Most organisations still have no AI policy

38%

of organisations have a formal AI policy.
A quarter have no active policy at all

6%

of boards have a formal policy covering the board's own use of AI

82% of directors have used AI in board work in the last six months.

ISACA AI Pulse Poll 2026, May 2026, 3,400+ digital trust professionals (38%) | Diligent Director Confidence Index, June 2026, 104 US public company directors (6% | 82%)

2026

What’s changed this year

148

AI incidents logged in a single quarter.
Agentic AI, cyber and operational misuse
is now a category of its own

116

companies signed a joint warning
in August of a limited window
to strengthen defences

Attacks now complete in minutes. Response plans still count escalation in hours.

AI Incident Database, May to July 2026 (148 incidents) | Joint industry letter, 27 August 2026 (116 signatories)

WHERE THE TWO MEET

A wide range of AI risks. Some land on your cyber perimeter.

What security leaders say they are least prepared for

50%

Attacks on their own AI systems

33%

Ransomware

PwC Global Digital Trust Insights 2027, October 2026, n=3,934 executives across 71 countries (50% | 33%)

RISK ONE

Shadow AI

It is not malicious. It is a high performer.

57%

of employees hide their AI use, and present AI-generated work as their own

46%

have put company information or IP into public AI tools

43%

of security incidents involved shadow AI. Double the year before

Of the organisations attacked through their own AI models, 92% had not properly controlled access to them.

Trust, attitudes and use of AI, Melbourne Business School with KPMG, 2025 (57% | 46%) | IBM Cost of a Data Breach 2026 (43% | 92%)

RISK TWO

Agentic AI

AI that acts, not just answers

21%

have a mature governance model for agentic AI

80%

of the Fortune 500 run AI agents built with low-code tools

AWS, December 2025

A human gave an AI coding agent their own access level. That bypassed the two-engineer sign-off required for production changes.

The agent deleted and recreated a live environment. 13-hour outage.

And the other direction: in the last fortnight JPMorgan, Citi, Bank of America and Schwab all moved to block third-party AI agents from customer accounts.

Deloitte State of AI in the Enterprise 2026, 3,000+ leaders across 24 countries (21%) | Microsoft, reported by Orro Group, October 2026 (80%) | JPMorgan Q3 earnings call, 14 October 2026, and Financial Times, 19 October 2026

RISK THREE

Prompt injection

Social engineering, aimed at the AI

An instruction hidden in a document, an email, or a skill your team downloaded to work faster. Your AI reads it and obeys.

Your people are installing these themselves, in seconds, to work faster. Often without procurement or security review, and no easy way to see what is inside.

OpenAI has said publicly that prompt injection is unlikely ever to be fully solved. This is managed, not fixed.

Snyk ToxicSkills research, February 2026

THE REAL GAP

Adoption moves in days. Governance takes months.

1 billion

people use ChatGPT every week.

47%

say their organisation has previously not followed its own AI governance process for urgent deployments

Ask for an interim position this quarter, and an amnesty to find out what is actually in use.

OpenAI, 6 August 2026 (1 billion weekly users) | EY US AI Risk and Governance Survey, September 2026, 202 senior AI decision-makers at US listed companies above $1bn revenue (47%)

Five questions for your board

1	**Appetite**	What is our risk appetite for AI use, and has our risk appetite statement been updated to reflect it?
2	**Visibility**	How are we tracking AI use across the organisation, including public tools on personal accounts?
3	**Authority**	Which AI systems hold access to our systems, who authorised that access, and who can switch them off?
4	**Assessment**	What are we using to tell high-risk AI applications from low-risk ones?
5	**Resilience**	If an AI-enabled incident disrupts a critical service, who has the authority to shut it down, and how do we keep operating once they do?