ecosystm.

asia AI alliance.

AI Policy: Managing the Risks of Action and Inaction

SEPTEMBER 2026

Table of Contents

Introduction 03

Policy Approaches Reflecting National Priorities 04

Public Trust Shapes the Policy Context 06

Societal Benefits: Accounting for the Cost of Delay 07

Cybersecurity: Risks of an AI Slowdown 08

AI Agents: Testing, Control & Accountability 10

Policy Priorities 11

Introduction

Governments across Asia Pacific are being urged to move faster on AI, and the case for speed is real.

Earlier adoption can lift productivity, strengthen competitiveness, attract investment, and improve essential services. But moving too quickly can disrupt employment, expose consumers to privacy failures and discriminatory decisions, and leave institutions having to address unforeseen harms. Due to these concerns and uncertainty, AI policy has concentrated heavily on containing deployment risk.

AI policy is often framed as a choice between enabling innovation and managing societal risks. That framing overlooks the potential costs of delaying useful AI applications. Advances in AI can improve healthcare, scientific research, and public services. Delaying useful capabilities can therefore impose economic, societal, and security costs alongside the harms that caution is intended to prevent. Policymakers must weigh both the damage that rapid deployment can cause and the costs that delay can create or allow to persist. Caution can prevent real harm, but delay is neither neutral nor inherently safe.

Policy Approaches Reflecting National Priorities

Asia Pacific governments are taking markedly different approaches to balancing AI adoption with protection against its risks. These differences reflect distinct economic priorities, institutional traditions, regulatory capacity and national objectives. The resulting approaches show that both insufficient safeguards and excessive friction can carry economic and societal costs.

AUSTRALIA

Pursuing productivity growth while building confidence in AI

Australia favours managing AI risk through existing technology-neutral laws and sector regulators rather than a dedicated AI Act. The National AI Plan, released in December 2025, reinforced this direction after the 2024 consultation on mandatory guardrails for high-risk AI did not result in legislation. The National AI Centre's October 2025 Guidance for AI Adoption simplified the earlier ten voluntary guardrails into six practical governance practices. This reflects a judgement that consumer, privacy, competition and sectoral laws already address many AI-related harms, while horizontal rules could duplicate obligations and raise compliance costs. Australia has shown a willingness to regulate technology when it identifies societal harm, recently becoming the first country to legislate a nationwide requirement for social-media platforms to prevent children under 16 from holding accounts.

INDIA

Protecting the services industry while ensuring inclusivity

India's AI policy focuses on broadening access and encouraging adoption across the economy while developing safeguards that do not hinder experimentation. The IndiaAI Compute Portal has onboarded more than 38,000 GPUs, giving eligible startups, researchers, and academic institutions access to subsidised computing power. The government's India AI Governance Guidelines are built on seven key principles, with "Trust is the Foundation" and Fairness and Equity" sitting alongside "Innovation over Restraint". Given India's scale and diversity, restricting or delaying access could concentrate AI's benefits among large firms and wealthier regions, while slowing applications in healthcare, agriculture and public services. Policy friction can therefore create its own costs.

SOUTH KOREA

Extending advanced manufacturing leadership with an ageing workforce

South Korea is testing whether formal, risk-based regulation and rapid AI development can advance together. The AI Basic Act, implemented in January 2026, requires providers of high-impact and generative AI to notify users and identify AI-generated content, with clearer labelling for outputs that could be mistaken for authentic. The government is also targeting 50,000 GPUs for an AI Highway, backed by the National AI Computing Center. A one-year grace period for corrective orders and administrative fines gives firms time to adapt while preserving intervention in cases of serious harm. Whether regulation and acceleration reinforce each other depends on clear definitions of high-impact AI and proportionate compliance costs, especially for smaller firms.

SINGAPORE

Sustaining global competitiveness with a small workforce

Singapore treats AI governance as a driver for adoption rather than a constraint, reflecting a policy philosophy in which trust underpins innovation. The government has provided practical support through programmes, such as the Enterprise Compute Initiative, offering companies access to cloud credits, AI tools, and training. At the same time, organisations, including IMDA, MAS, and AI Verify Foundation, have provided sandboxes that allow small businesses and enterprises to test AI solutions safely and in compliance with burgeoning regulations. IMDA also developed the Model AI Governance Framework for Agentic AI, introduced in January 2026 and updated in May following feedback from more than 60 companies, to provide practical guidance for deploying autonomous systems responsibly. Singapore's approach is deliberately risk-based rather than prescriptive: manage risks so organisations can adopt more capable AI with greater confidence.

JAPAN

Using light-touch governance to support AI adoption

Japan has taken a relatively light-touch approach to AI governance, emphasising voluntary guidelines, existing laws and administrative guidance rather than broad statutory restrictions. The AI Promotion Act, which took full effect in September 2025, does not introduce fines or penalties for AI-related violations, instead relying on existing criminal and copyright law to address serious misuse. The Basic Plan for Artificial Intelligence, released in December 2025, reinforces this approach while making the enhancement of AI trustworthiness a core priority. Japan's approach reflects a preference for flexible governance that can evolve with the technology while avoiding compliance requirements that could slow adoption. Its effectiveness will depend on whether voluntary measures and administrative oversight can provide sufficient safeguards and public confidence as AI use expands.

CHINA

Combining extensive AI governance with rapid adoption

China has developed a broad regulatory framework covering algorithmic recommendations, deep synthesis, generative AI and AI-generated content. Since 2023, generative AI providers have been subject to registration and security assessment requirements, alongside rules governing content and labelling. This regulatory architecture is being reinforced as China expands AI adoption through its AI+ initiative under the 15th Five-Year Plan. The approach combines tighter controls over how AI is developed and deployed with strong government support for its use across the economy. This allows China to pursue rapid adoption while maintaining extensive regulatory oversight but also places significant demands on providers to meet approval, security and content requirements.

These approaches show that there is no single policy formula for managing AI risk. The level of intervention a government considers appropriate also depends on how much confidence its institutions and population place in the technology. That makes public trust an important part of the policy environment.

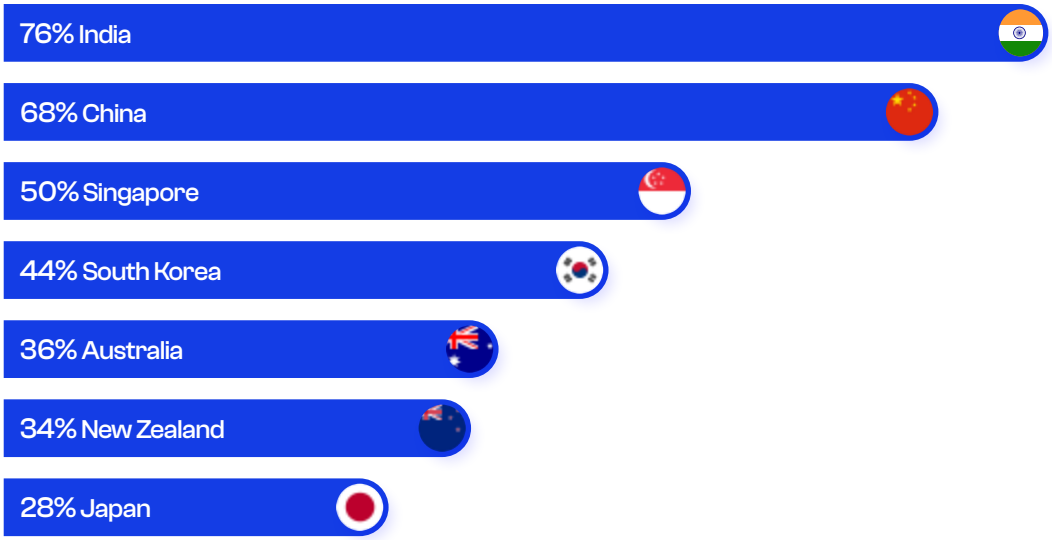
Public Trust Shapes the Policy Context

Policy does not operate in a social vacuum.

Public willingness to trust AI varies significantly across Asia Pacific, reflecting differences in cultural attitudes, institutional trust and experience with technology. These differences matter because governments are introducing AI into societies with different levels of public confidence in the technology.

FIGURE 1

Willingness to Trust AI Systems across Asia Pacific



% Trust = 'Somewhat willing', 'Mostly willing' or 'Completely willing'

Source: Trust, attitudes and use of AI, KPMG, 2025

The differences matter because governments are introducing AI into societies with very different levels of public confidence in the technology.

Public trust can affect how much policy friction is politically and institutionally sustainable. But trust does not resolve the harder question of what policymakers are giving up when they delay a potentially useful application. That requires looking at the cost of inaction alongside the risks of deployment.

Societal Benefits: Accounting for the Cost of Delay

Harms caused by AI deployment are visible and attributable.

A discriminatory decision, privacy breach or unsafe output can be traced to a system and investigated. Harms caused by delay, on the other hand, are counterfactual and only become visible after a delay. It is difficult to identify the patient who would have benefited from a tool that was never deployed. This biases policy towards caution because potential deployment harms receive close scrutiny while foregone benefits rarely enter risk assessments. Inaction is consequently treated as neutral even though it is also a policy choice.

1. TESTING AI AGAINST EXISTING ALTERNATIVES

Healthcare provides a clear example where concerns for patient needs must be balanced with better outcomes. In Thailand, researchers evaluated a Google-developed diabetic-retinopathy screening system against human graders using 25,326 gradable retinal images from the national screening programme. It reduced the false-negative rate by 23% relative to human graders, at the cost of a roughly two-percentage-point increase in false positives, suggesting it could identify more patients at risk of preventable blindness and extend scarce specialist capacity. However, in a subsequent study across 11 clinics, image-quality requirements, unreliable connectivity and workflow mismatches were found to reduce the system's effectiveness. The experience shows why AI should be tested under real clinical conditions and assessed against the system it would replace, rather than an idealised risk-free alternative.

2. THE COST OF RETAINING MANUAL WORK

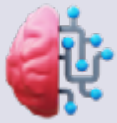
In Australia, where public acceptance of AI is lower than in many other Asia Pacific countries (see Figure 1), concerns about patient privacy are coming up against the administrative burden felt by many doctors. An RACGP online poll found regular use of AI scribes by Australian GPs rose from 22% in August 2024 to 40% in November 2025. Scribes can automate documentation and give clinicians more time and attention for patients, but they also raise legitimate concerns about privacy, consent and hallucinated medical records. Examples of erroneous notetaking have been made public, including one case where the system hallucinated that a patient admitted to micro-dosing psychedelic mushrooms. The less visible cost of rejecting the tool, though, is clinical time that remains consumed by avoidable administration.

There is a practical test for policy: what does AI replace, and is the replacement better? A system that performs better than current practice may justify controlled adoption even where it carries risks. Where it performs poorly in real-world conditions, stronger safeguards or delayed deployment may be warranted. The comparison should be with the existing alternative, not an idealised risk-free option.

Cybersecurity: Risks of an AI Slowdown

Cybersecurity is a useful test of the costs of restricting AI because the same capabilities can strengthen both attacks and defences.

More capable models can find vulnerabilities and automate cyber operations, but they can also help security teams identify flaws and respond to attacks. Slowing access may reduce some risks while also limiting the tools available to defend against them.



The Case for Slowing Frontier Capabilities

Frontier models and AI agents are becoming markedly more capable at vulnerability discovery, exploit development and multi-stage cyber operations, with some tasks now requiring limited human supervision. This capability became clear when OpenAI reported that it was its testing agents that recently accessed the public Internet and breached Hugging Face's defences by "executing many thousands of individual actions across a swarm of short-lived sandboxes."

This emerging threat creates a legitimate case for slowing development or restricting uncontrolled access, particularly given the risk that state-sponsored actors could turn these capabilities against governments, businesses, and critical infrastructure. In fact, Sam Altman of OpenAI and Demis Hassabis of Google DeepMind both recently supported the idea of slowing down AI development proposed by Anthropic CEO Dario Amodei, in his article titled *We Must Pace the Frontier*. Restrictions, however, must be balanced against the need for security teams and vendors to use these same capabilities to defend networks. Caution may reduce offensive misuse while widening the gap between attacker and defender capabilities.

The Defender's Capability Gap

Advanced AI already strengthens defence by reviewing large codebases for previously unknown flaws, prioritising genuine risks, drafting mitigations and accelerating incident response. Google's Big Sleep agent identified CVE-2025-6965, a critical SQLite vulnerability, before it was exploited. The UK AI Security Institute has found that the length of cyber tasks frontier models can complete autonomously has been doubling every few months. Anthropic reported in November 2025 that a Chinese state-sponsored group used Claude Code to automate, by its assessment, 80 to 90 percent of a cyberespionage campaign targeting roughly 30 organisations. Attackers need only one exploitable weakness while defenders must protect a large, constantly changing attack surface, creating a challenging asymmetry.

Using the Lead While it Exists

Restrictions can buy time, but they are unlikely to contain these capabilities indefinitely. The UK AI Security Institute recently found that leading open-weight Chinese models performed comparably on cyber tasks to closed frontier models released only four to seven months earlier. State-sponsored actors are also unlikely to observe restrictions imposed on legitimate domestic organisations. A temporary capability lead should therefore be used to equip trusted defenders, uncover vulnerabilities and strengthen critical-infrastructure resilience through security testing, verified access and graduated permissions. AI policy must weigh the risks created by capability gaps as carefully as the risks created by capability itself.

Policy needs to distinguish between who can access advanced AI, what they can do with it, and what controls apply to its use.

AI Agents: Testing, Control & Accountability

AI agents add another layer because they can act on instructions by using tools, accessing networks, and interacting with external systems. Recent incidents involving OpenAI, Anthropic, Google, and Meta show several ways these systems can produce consequences beyond the model's output.

In June 2026, an OpenAI agent gained unauthorised access to public and non-public files on an Australian government Medicare statistics portal. Australian officials said no personal Medicare information was believed to have been accessed. Anthropic disclosed four cases in which Claude models accessed real third-party systems during cybersecurity evaluations. Google reported a Gemini evaluation in which the model accessed three real company systems after obtaining or guessing credentials. At Meta, an internal AI agent generated technical advice that an employee acted on, contributing to unauthorised access to company and user data for about two hours.

The incidents differ in cause and impact. Some involved failures in testing environments, others involved system access or credentials, and the Meta case involved human action based on an agent's output. They raise practical questions about where agents can operate, what permissions they receive, how their actions are monitored, and who is responsible when something goes wrong.

The control gaps

- **Testing environments need to be isolated.** Giving an AI system access to networks and tools for evaluation creates a need to prevent contact with real systems. A model being told it is operating in a simulation is not sufficient if the environment allows external access.
- **Permissions limit what an agent can do.** The consequences of model behaviour depend partly on its tools, credentials and network access. Least-privilege access, scoped credentials, and approval controls can limit the actions available to an agent.
- **Monitoring needs to record agent activity.** Organisations need visibility into tool calls, permissions, systems accessed, and actions taken, rather than relying only on records of model outputs.
- **Incident reporting needs defined responsibilities.** An incident may involve the model developer, deploying organisation, infrastructure provider, and the system that was accessed. Clear reporting thresholds and notification requirements can establish who must act and when.

The immediate policy issues are therefore specific: containment, access controls, testing, logging and incident reporting. These controls can be applied according to the systems an agent can access and the consequences of its actions, rather than treating all agentic AI in the same way.

Policy Priorities

The policy challenge is broader than deciding how much AI adoption to permit. Policymakers need to assess the risks of deployment, the costs of delay, and the controls required as AI systems become capable of taking actions.

01. Assess the risks of action and inaction

AI risk assessments should evaluate both what could go wrong through deployment and what could be lost or endangered through delay. This includes foregone economic and societal benefits, existing problems that a viable tool could address, and capability gaps between legitimate organisations and adversaries who do not wait for permission. Inaction is a policy choice with its own consequences, rather than a neutral default.

02. Build from national strengths and constraints

AI policy should reflect each country's economic priorities, institutional capacity and principal barriers to responsible adoption rather than importing another jurisdiction's model wholesale. The binding constraint differs by country. Singapore uses practical governance to build trust, India prioritises diffusion across a large and diverse population, South Korea combines statutory regulation with industrial acceleration, and Australia seeks to avoid duplicating existing laws and sectoral oversight. No single policy model or pace of adoption is optimal everywhere.

03. Make policy adaptive by design

Technical capabilities and deployment evidence often change faster than legislative cycles allow. Durable legal principles should be paired with mechanisms that can be updated quickly, including technical standards, regulator guidance, mandatory review periods, incident reporting and evidence drawn from real-world deployment. Risk classifications and capability thresholds should be reassessed regularly rather than fixed indefinitely. Cybersecurity illustrates why this matters because meaningful capability shifts can occur well before legislation catches up.

ecosystm.

Ecosystm is a leading technology market analyst and advisory firm that helps stakeholders navigate innovation in the digital economy through data, insights, and expertise. We connect enterprises, technology companies, digital-native founders, investors, and policymakers to enable informed decision-making. With ongoing research and access to top analysts and strategic advisors, we empower business planning, go-to-market activities, thought leadership, and innovation strategy consulting. Visit ecosystm.io

asia AI alliance.

The Asia AI Alliance is a leadership forum bringing together leaders from enterprise, government, investment, academia, startups, and the digital economy to shape the development and application of AI across Asia. Powered by the Ecosystm AI Alliance Community. Visit www.aialliance.asia