EINPRESSWIRE

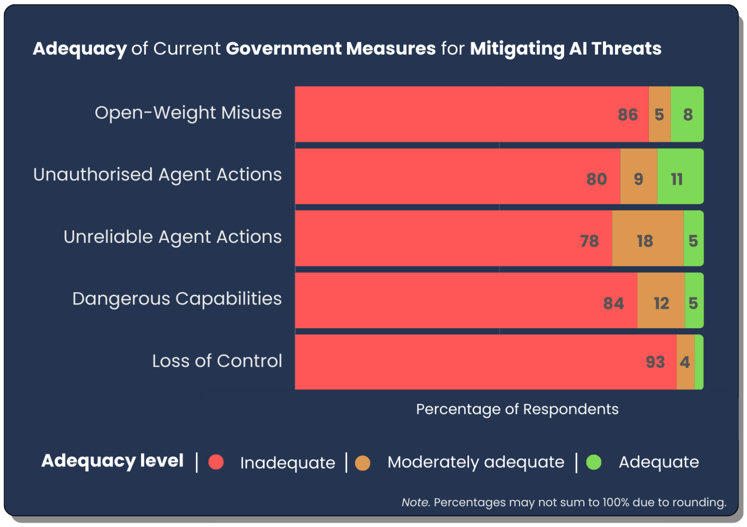
As Government Splits on AI Rules, New Report Reveals Critical Gaps

Expert stress test provides new evidence as Labor splits on approach to AI risks

CANBERRA, AUSTRALIA, August 19, 2025 /EINPresswire.com/ -- As the Albanese government remains divided on AI regulation, a new expert report reveals the risks from general-purpose AI models that no regulator can address alone.

The AI Legislation Stress Test answers the Government's call for a gap analysis to inform next steps on AI regulation. The stress test drew on 64 experts in AI and public policy from top universities, law firms and AI labs.

The vast majority of experts found current laws inadequate across all five AI threats they evaluated.



The vast majority of experts found current laws inadequate across all five AI threats they evaluated.

Australians For AI Safety

Australians for AI Safety logo

“The key takeaway from the report is that no regulator is addressing the risks of general-purpose AI, and many of those risks are serious,” said Greg Sadler, the report’s lead author. “No one is disputing that the TGA is best placed to manage AI in medical devices or that CASA is best placed to manage AI in aviation, but the experts found that no one is in charge of these broader concerns.”

Experts found that 4 of the 5 AI threats they assessed were a “realistic probability” or “likely” to lead to harm in Australia within the next 5 years.

“AI is a complex supply chain of developers, deployers and users. Effective regulation has to match the risk to the person who can best manage it,” said Greg Sadler “Some AI risks come from 'black boxes' made by the developer who built it, and there's not much Australian business can do to manage those risks. Making Australian businesses responsible for risks they can't manage is like telling Australians that they have to install their own airbags and crumple zones in their cars."

Experts ranked AI models giving people "access to dangerous capabilities" as their top policy priority. Access to dangerous capabilities refers to AI models that could help people conduct cyber attacks or even build bioweapons.

"AI labs have already released detailed reports of malicious actors using their products to assist with cyber attacks. These risks are real," said Justin Olive, the Head of AI safety at Arcadia Impact, who contributed to the report.

Experts assessed the consequence of bad actors using advanced AI to access dangerous capabilities as "severe", meaning it would kill between 200-1000 Australians per year or cause $2-$20 billion in economic harm. 42% of experts went further, assessing the risk as "catastrophic", estimating it would kill over 1000 Australians per year or lead to over $20 billion in economic damages unless Government puts effective regulation in place.

AI labs have been raising the alarm about the risks of their models being misused to launch cyber attacks or build bioweapons. OpenAI said GPT-5 might be able to provide meaningful counterfactual assistance to "novice" actors that enables them to create known biological or chemical threats. OpenAI says this leads to a significantly increased likelihood and frequency of biological or chemical terror events by non-state actors.

While OpenAI claims to have voluntarily put safeguards in place for GPT5, an evaluation by the UK's AI Security Institute "identified multiple model-level jailbreaks that overcome GPT-5's built-in refusal logic." OpenAI acknowledges that "there is a risk of previously unknown universal jailbreaks being discovered after deployment."

The report suggests the voluntary safeguards are not appropriate for such severe risks.

On 1 August 2025, Google issued a similar warning about the chemical, biological, radiological, and nuclear (CBRN) threats of Gemini 2.5 Deep Think. Google said "the model has enough technical knowledge in certain CBRN scenarios and stages to be considered at early alert threshold". This threshold is defined as the model's ability to "significantly assist a low-resourced actor with dual-use scientific protocols, resulting in a substantial increase in ability to cause a mass casualty event".

The report found that an Australian AI Act is best placed to require AI developers to assess the risks of their models and ensure adequate safeguards are in place before the models are deployed to the public.

The report also highlighted risks of “AI agents”. AI developers are building agents designed to complete online tasks over extended periods autonomously.

“Agents are hugely unreliable, hackable, difficult to track, and cheap to use. They are a gift to grifters and other unscrupulous people,” said Jamie Freestone, a Philosopher at the Australian National University, who contributed to the report.

The report found that AI agents can cause harm in two different ways. Users could rely on AI agents that are not competent and engage in behaviours like deception, fabrication, and hallucination. Alternatively, users could direct an AI agent towards one goal, but the agent autonomously pursues other goals. An AI agent could potentially commit crimes that the user didn’t authorise without the agent’s creator being held responsible.

"We're building increasingly autonomous AI systems without the governance infrastructure to manage containment failures. That's a dangerous gamble with potentially catastrophic stakes," said Alexander Saeri, an AI Governance Researcher MIT FutureTech, who contributed to the report.

The report comes after the Productivity Commission recommended a pause on “mandatory guardrails” and Labor's caucus splits between Industry Minister Tim Ayres' "lighter touch" approach and former minister Ed Husic's calls for a comprehensive framework.

"While politicians debate, AI capabilities are advancing rapidly. The report provides concrete evidence about which approach would actually work," said Luke Freeman, the report’s co-author.

The analysis is timely given the ongoing Economic Reform Roundtable, where AI regulation will be a flashpoint. The report supports the finding of Ed Husic's two-year consultation conclusion that "we need an economy-wide approach to a technology that will touch every corner of the economy."

However, the report also shows Minister Ayres' existing regulatory approach is relevant to addressing some risks, if properly resourced and coordinated.

The new report suggests both camps are partially right: strengthen existing regulations where possible, but acknowledge that general-purpose AI has some novel risks that require new measures.

The full report is available for download at https://www.goodancestors.org.au/ai-stress-test

Mr Gregory Sadler
Good Ancestors Policy
+61 401 534 879

email us here

This press release can be viewed online at: https://www.einpresswire.com/article/840781277