

Background Note #1¹

Navigating the AI Policy Crisis

Australian AI policy faces five immutable realities:

- AI is already having radical impacts on Australia's national security, its counter-intelligence and the human security of its citizens. These impacts will continue to trend towards greater destabilisation of international order and degradation of the social order inside the country.
- The most transformative negative impacts of AI on Australia will be delivered by systems owned and operated by extremely powerful actors based in the United States and China. At the same time, less advanced AI systems developed in Australia or other partner countries, have the potential for profound positive impacts on the national interest.
- Navigating access to benefits in that ecosystem, like ensuring protection from misuse and hostile intent, will be most effective for Australia when the country leads international collaborations to those ends, both with the US and China, but also with other middle powers and multilateral initiatives. The European Union, the UK, France, Germany, Canada and Singapore must become first rank AI partners for Australia.
- Central dilemmas in AI policy are already highly politicised through direct interventions and dispositions of a technology oligarchy often loosely associating around anti-democratic dispositions.
- Australia cannot wait until international AI regimes settle into place since high levels of uncertainty remain: the country will need to be at the leading edge of experimentation and risk-taking among its peers for years to come. Micro and meso level actions (operational) will be essential before the regimes at the macro (strategic) level stabilise.

We need to be deliberate in calling policy making for AI as being in "crisis" rather than seeing it merely as a collection of difficult policy problems that can be sequentially addressed. It is a crisis on at least two levels: domestic policy and international relations. There are three defining criteria of a crisis: an existential threat, accelerating urgency, and profound uncertainty amidst fierce military and technological international competition among nuclear-armed great powers.

There are over-arching reasons why this framing as a "crisis" is more than academic.

First, a country must ensure its policies are holistically framed where there is explicit recognition at all levels of policy-making that actions on one level or in a narrow area of policy can influence other levels.

Second, defining the challenge as a crisis is important because we need to understand that the dynamics of decision-making for a crisis can be different for what might be called normal times. Above all, crises can erode accountability for the power-holders and undermine the democratic political influence of the stakeholders, both citizens and businesses.

¹ Prepared as background for a submission by the Social Cyber Institute to the Joint Select Committee on Artificial Intelligence on 14 September 2026, https://www.apf.gov.au/Parliamentary_Business/Committees/Joint/Artificial_Intelligence/ArtificialIntelligence.

We can ground the "crisis" label in the pattern of warnings already being issued by senior figures, while noting the absence of any matching institutional response: warnings have been coming thick and fast, from people like the UN Secretary-General, the Pope and other religious leaders, world-leading scientists and Nobel laureates, and leaders of frontier AI companies, while the governments involved and the corporations themselves have yet to agree coherent, comprehensive and enforceable safeguards, or even basic standards. In plain terms: AI systems are moving faster than governance frameworks and political decision-making. The leading AI corporations are simultaneously involved in stimulating economic gains and supporting military, intelligence and security confrontations among great powers in the midst of an increasingly destabilised global system. The policy gaps are creating uncertainty and volatility. Responsibilities are split across institutions. The result is uneven risk management, weak accountability, lack of standardisation, and diminished positive influence of actors who are outside the AI corporate giants.

Crisis decision-making: the four Cs

The analytical centrepiece of this submission is to accept "crisis decision-making" as a distinct field of social science, and to treat it as the absolute central focus for anyone wanting to influence the direction and outcomes of the AI Crisis. We can root this in a substantial literature, with hundreds of books on it, on everything from the Cuban Missile Crisis in 1962 to the climate emergency we face today, including, of course, the crisis decision-making around the COVID pandemic. We draw in particular on a study of EU crisis-mode decision-making. The core claim from this literature is blunt. It is widely accepted in the social sciences that a severe crisis distorts decision-making so much that new mechanisms and modes of political decision-making must be put in place and these must reflect the unique characteristics of the crisis at hand.²

To structure our approach to the current AI crisis response we reference a 2020 study³ which drew on the Covid experience and postulated four Cs of crisis decision-making: cognition, communication, coordination and control.

On *cognition*, there is no consensus on AI policy, globally or in Australia. Not only is there no agreement on what constitutes safe AI; there is no agreement on whether we face a crisis at all, or what kind. The UN Secretary-General, religious leaders, Nobel laureates, senior AI scientists and heads of government have all warned of catastrophic or existential risk, yet almost no government has adopted that framing formally, let alone legislated on it. A third set of actors complicates the picture further. These are the executives of the major AI firms, who are conducting their own contest over AI policy largely outside democratic processes.

On *communication*, our diagnosis is that governments are sending incoherent signals not simply because they are confused, but because the messages serve different purposes. The same government will warn of existential risk in one forum and, in another, describe regulation as a threat to national competitiveness, while imposing weak rules and claiming they are safeguards. The pattern in

² Asthana, S., et al. "Governance and Public Health Decision-Making During the COVID-19 Pandemic: A Scoping Review." *Public Health Reviews*, vol. 45, 2024, doi:10.3389/phrs.2024.1606095, <https://www.ssph-journal.org/journals/public-health-reviews/articles/10.3389/phrs.2024.1606095/full>; Boin, Arjen, et al. "Hybrid Crisis Governance: A Systematic Review of Control-Based and Capacity-Based Approaches." *Risk, Hazards & Crisis in Public Policy*, vol. 17, no. 1, 2026, pp. 7–38, doi:10.1002/rhc3.70054, <https://onlinelibrary.wiley.com/doi/10.1002/rhc3.70054>; Nikolić, Marina. "Crisis Windows and Reversible Empowerment in the EU's Foreign Policy." *Journal of European Public Policy*, 2026, doi:10.1080/13501763.2026.2665384, <https://www.tandfonline.com/doi/full/10.1080/13501763.2026.2665384>.

³ Lousie Comfort, Naim Kapucu, Kilkon Ko, Scira Menoni and Michael D. Siciliano (2020). Crisis Decision-Making on a Global Scale: Transition from Cognition to Collective Action under Threat of COVID-19. *Public Administration Review*. 80. 616-622. 10.1111/puar.13252.

the policy itself is consistent with this: governments are responding to discrete harms (deepfakes, fraud, misuse of personal data) with piecemeal measures that are often reasonable on their own terms, while declining to act on, or even acknowledge, the systemic challenge their own comments imply.

Coordination is where our language must be sharpest, because it is the defining failure of the international response. Not until July 2026 did the world's governments convene as a whole on AI: the Global Dialogue on AI Governance, mandated by General Assembly resolution A/RES/79/325 and held in Geneva. Everything before it (Bletchley, Seoul, Paris) was invitation-based and excluded most of the world from decisions that will affect all of it. Even the Dialogue's mandate dates only to August 2025, years into the deployment of frontier systems. In the interim, the world leaned on AI Safety Institutes as the vehicle for policy. That reliance was misplaced, and predictably so. Safety institutes evaluate; they do not bind. Establishing one lets a government demonstrate seriousness without accepting any constraint on the firms it is meant to be governing, a process that allows measurement to substitute for regulation. Likewise, an Office of AI is an internal administrative group with good intentions but no authority to ensure consultation and co-ordination across key stakeholders under the influence of legislated principles and standards.

What does *Control* look like during the AI policy crisis? The picture is equally fragmented and, in important respects, deteriorating. Control in crisis decision-making refers to the capacity of political authorities to set binding rules, enforce compliance, and align the behaviour of powerful actors with collectively agreed objectives during a period of acute uncertainty and volatility. In the AI crisis, formal control remains weak: there is no global regulatory regime, no binding international treaty on frontier AI, and no agreed mechanism for overriding commercial incentives when they conflict with public safety. National and regional initiatives, such as the EU AI Act, emerging US executive measures, and assorted voluntary commitments by leading AI firms, are important but partial and fragmented. They operate at different speeds and levels of ambition, and with variegated policy ambitions. This produces a patchwork in which the most capable private actors can often choose their regulatory venue, delay implementation, or shape less than rigorous standards through lobbying. The result is a system in which control is effectively outsourced to a small number of corporations with their own self-serving agendas, while democratic institutions struggle to catch up, or even to understand the implications. Until governments can establish credible, enforceable constraints on the development and deployment of the most powerful AI systems, the “control” dimension of crisis decision-making will remain the weakest link in the global response.

If governments can stand up and say, “I’ve got an AI Safety Institute, I believe in responsible AI, I’ve got a framework for responsible AI,” then they think they’re doing something about the AI crisis. But in fact they’re not doing enough. Most AI safety institutes are not sufficiently funded to even begin to document the sorts of challenges presented by the AI crisis. This scepticism is particularly relevant to frontier AI regulation. It is the experience of AI Safety Institute people in the United States that commitments to AI safety made by private sector vendors of AI are not reliable and have been consistently flouted.

Australia as a case study of AI crisis decision-making

Australia to date provides a case study in the governance failures described above. The criticism is deliberately sharp, and it is directed at a consistent pattern rather than at isolated decisions: in each instance, the Australian government has adopted the language of AI risk while declining to accept any binding constraint arising from it.

The clearest illustration is the data centre build-out. On 23 April 2026, Microsoft announced a A\$25 billion investment in Australian AI and cloud infrastructure through to 2029, accompanied by a memorandum of understanding committing the company to the government's Expectations for Data Centres and AI Infrastructure Developers, published a month earlier. Those expectations (covering national interest, clean energy, water use, skills and local research) are not legislated. They create no obligation, carry no penalty, and are enforced by nothing beyond the signatory's own undertaking. Australia ranked second only to the United States in global data centre investment in 2024, and the government has described its regulatory environment as "rigorous but tech-friendly." The second half of that formulation is the operative one.

The contrast with jurisdictions attempting binding rules is instructive. Minnesota's 2025 data centre legislation (H.F. 16 / S.S.H.F. 16), enacted in June 2025, imposes environmental and energy requirements on large-scale facilities and empowers the Public Utilities Commission (PUC) over very large customers. The law establishes annual fees of \$2–5 million linked to peak electricity demand, requires pre-application evaluation and permit conditions for projects using more than 100 million gallons of water per year, allows the PUC to approve, modify or reject tariffs or energy supply agreements with data centres, requires each public utility to offer a clean energy and capacity tariff, includes large-scale data centres in the state's solar energy standard, and mandates green building or sustainable design certification within three years of operation. In Illinois, the POWER Act would have required data centres to fund their own generation and source it from renewables, submit quarterly water use reports, and meet water efficiency criteria, but failed to pass before the legislature's 31 May 2026 deadline. Instead, Governor JB Pritzker directed a two-year suspension of state tax incentives under the Data Center Investment Program, effective 1 July 2026, halting new agreements while protecting existing deals. Australia's expectations framework does not approach either the obligations contemplated in Illinois or those now operating in Minnesota.

The same pattern appears in Australia's position on lethal autonomous weapons, where the gap between accepting an argument and acting on it is unusually visible. Australia has participated in every Convention on Certain Conventional Weapons (CCW) meeting on autonomous weapons since 2016. At the 2021 Group of Governmental Experts (GGE) it stated that it saw no need at that time for a new legal instrument, and it has not since supported negotiating one. Yet at the 2023 GGE Australia endorsed the two-tier approach (that some uses of autonomous weapons are inherently unlawful while others are subject to limits under international humanitarian law) which is the framework underpinning the Chair's rolling text. Australia therefore accepts the substance while declining the procedural step. In September 2025, 42 states, roughly a third of the CCW's 128 high contracting parties, declared through a joint statement delivered by Brazil that the rolling text was a coherent basis for an instrument and that they were ready to negotiate. Australia was not among them, preferring a consensus-bound forum in which the states least interested in constraint hold an effective veto.

This is not yet the conduct of a government that believes it faces an existential risk, and it will shortly be tested. The GGE's three-year mandate concludes in 2026, and the Seventh Review Conference of the CCW meets in Geneva from 16 to 20 November 2026 to decide whether to

adopt a negotiating mandate, extend the discussion-only mandate, or let the process lapse. Australia's vote there will be a plainer statement of its position than any ministerial speech.

The Prime Minister's statement of 15 July 2026, given the title “AI in Australia’s interests”, completes the picture.⁴ It is framed overwhelmingly around economic opportunity and productivity. It claims, “a world-leading artificial intelligence framework to ensure Australia can capture the opportunity, share the benefits and keep Australians safe”. He goes on to say, “We will do that the Australian way, without undercutting conditions, dividing communities or damaging our environment.” The statement does not acknowledge the scale of the stakes its own government's rhetoric elsewhere implies, offers no substantive enforceable path forward, and makes no serious mention of controlling catastrophic risk. Taken with the Microsoft arrangement and the autonomous weapons record, it suggests that Australia's AI policy to date is an investment-attraction strategy with a safety vocabulary attached.

Dr Greg Austin
Co-founder and Research
Director
Social Cyber Institute

Dr Gary Waters
Distinguished Fellow
Social Cyber Institute

Prof Glenn Withers
Co-founder and Director
Social Cyber Institute

⁴ See “AI in Australia's interests”, Media release, Wednesday 15 July 2026
<https://www.pm.gov.au/media/ai-australias-interests>.